A Very Short Introduction

牛 津 通 识 读 本

网络

Detworks

[意大利] 圭多·卡尔达雷利　米凯莱·卡坦扎罗 / 著
李果 / 译

译林出版社

VISIT…

LANZAROTE Caliente.COM

A Very Short Introduction

牛 津 通 识 读 本

网络

Networks

[意大利] 圭多·卡尔达雷利　米凯莱·卡坦扎罗 / 著

李果 / 译

译林出版社

版权信息

Copyright © Guido Caldarelli and Michele Catanzaro, 2012
Networks was originally published in English in 2012.
This Bilingual Edition is published by arrangement with Oxford University Press
and is for sale in the People's Republic of China only, excluding Hong Kong SAR,
Macau SAR and Taiwan, and may not be bought for export therefrom.
Chinese and English edition copyright © 2018 by Yilin Press, Ltd
著作权合同登记号　图字：10-2012-485 号

书　　名　网络（牛津通识读本）

作　　者　【意】圭多·卡尔达雷利 米凯莱·卡坦扎罗

译　　者　李 果

责任编辑　何本国

出版发行　译林出版社

ISBN　9787544774383

关注我们的微博：@译林出版社

关注我们的微信：yilinpress

意见反馈：@你好小巴鱼

目录

CONTENTS

序言
第一章 从网络的观点看世界
第二章 富有成效的方法
第三章 网络世界
第四章 连接与闭合
第五章 超级连接器
第六章 网络的涌现
第七章 深入挖掘网络
第八章 网络中的完美风暴
第九章 整个世界是否就是一张网？
注释

献给我的家人

——圭多·卡尔达雷利

献给安娜

——米凯莱· 卡坦扎罗

序言

狄增如

著名诗人北岛有一首名为《生活》的一字诗：网。

毫无疑问，我们的世界运行在有形或无形的网络之中，尤其是在信息技术飞速发展的今天。进入21世纪以来，技术进步和社会经济的全球化，把世界更加紧密地联结在一起，系统性和复杂性已经成为各领域核心问题的共性特征。

正像《网络》一书中指出的，系统性和复杂性的现象在我们的生活中比比皆是，大到全球性的金融危机、生态系统的稳定性，小到空手道俱乐部的分裂和足球队的场上表现，其突出表现是系统各单元或要素，通过相互关联和影响所展现出的涌现性行为：在没有中心控制和全局信息的情况下，仅仅通过个体之间的局域相互作用，系统就可以在一定条件下展现出宏观的时空或功能结构，在新的层次上涌现出具有整体性的性质和功能。在这里应该强调的是，非线性的关联和相互作用关系对于系统的涌现性至关重要，它使得系统的整体行为不能通过个体行为的简单叠加而获得。因此，结构决定系统功能，刻画系统中的相互作用关系对于认识和理解系统的宏观性质意义重大。

而复杂网络就是对系统结构的最一般和本质的刻画。系统中的个体对应于网络的节点，个体之间的联系或相互作用对应于联结节点的边。网络的抽象，超越了每个系统自身的特殊性质、独特背景和演化机制，让我们可以用统一、普适的网络分析方法去研究系统的性质，进而加深对系统共性的了解。事实上，自18世纪欧拉建立起图与网络的概念，解决了哥尼斯堡七桥问题开始，图与网络就成为刻画结构的重要工具，并于19世纪中叶进入社会学领域，社会网络分析由此成为社会计量学的重要工具。

1998年，沃茨——斯托加茨提出小世界网络模型，指出少量的随机捷径会改变网络的拓扑结构，从而涌现出小世界的效应；随后，巴拉巴西提出了无标度网络概念，解释了增长和择优机制在复杂网络自组织演化过程中的普遍性和幂律的重要性。这些开创性的研究工作使我们认识到，实际系统的网络结构存在着许多超越经典随机网络的特

殊性质，例如小世界性质、幂律度分布、不同的匹配关系、社团结构等，我们需要建立更加细致、准确的概念，刻画网络结构并进而更加科学地认识系统性质，以促进网络科学的快速发展。

网络描述方法已被广泛应用于实际系统的研究，例如神经元网络、食物链网络、因特网络、赛博空间以及人与人之间交往的社会网络等等，研究结果加深了对这些具体系统的理解，并且提出了一系列新的概念和分析方法。总体上说，网络科学研究包括以下主要内容：一、如何定量刻画复杂网络？通过实证分析，了解实际网络结构的特点，并建立相应概念以刻画网络结构特征；二、网络是如何发展成现在这种结构的？建立网络演化模型，理解网络结构的产生和涌现；三、网络特定结构的后果是什么？利用基于网络的特定动力学过程，如新陈代谢网络上的物质流、食物链网络上的能量流、万维网上的信息传播、社会网络上的舆论形成等等，探讨网络结构与功能的关系；四、利用网络结构的设计和调整，控制和优化系统功能。

意大利的圭多·卡尔达雷利教授，是网络科学领域的知名专家，在复杂网络研究中，特别是社会、经济网络的分析和应用中颇有建树。《网络》一书，结合许多生动有趣的案例，如六度分隔实验、空手道俱乐部、万维网、食物链网络、基因调控网络等，深入浅出地介绍了网络科学的发展历程、核心概念和最新进展。书中没有一个数学公式，但通过具体案例的解读，能够让大家科学准确地把握网络分析的基本概念，如小世界性质、无标度网络的异质性、网络中的社团结构、无标度网络的脆弱性和鲁棒性、网络结构与传播的关系等，展现了作者强大的学术功力。可以说，《网络》一书，是大家了解和进入网络科学领域的非常好的一本入门读物。

总的来说，复杂系统的涌现现象是具有整体性和全局性的行为，不能通过分析还原的方法去研究，必须考虑个体之间的关联和相互作用。从这个意义上讲，理解复杂系统的行为应该从理解系统相互作用的网络结构开始。2009年，《科学》杂志以“复杂系统与网络”（Complex Systems and Networks）为主题，发表一集专刊，其中巴拉巴西教授的一段话很有启发意义。他指出，由于底层结构对于系统行为有着重大的影响，除非探讨网络结构，否则没有办法去理解复杂系统。希望《网络》一书能够带领大家进入生动有趣的网络科学领域。

第一章 从网络的观点看世界

网络存在于许多人的日常生活之中。通常的一天里，我们都会查收电子邮件，更新社交网络上的个人资料，打移动电话，使用公共交通工具，乘坐飞机，转移货币和货物，或者开始新的私人和业务关系等等。我们在所有这些情况中——有意或无意地——都会涉及网络及其特性。同样地，网络也会出现在重要的全球现象中。金融危机会在银行和公司的联系网络中产生多米诺效应。流行病，比如禽流感、非典型肺炎或者猪流感，则在机场网络之间蔓延。气候变化能够改变生态系统内部不同物种之间的关系网络，恐怖主义和战争则瞄准了一国之中的基础设施网络。电网内会发生大规模停电，计算机病毒在互联网中传播，政府和企业则会通过人们的社交网络和其他数字通信工具来追踪他们的身份信息。最后，遗传学的各种应用取决于人们对在细胞内部起作用的基因调控网络知识的掌握。

在所有这些情况中，我们都会处理大量不同元素（个人、公司、机场、物种、发电站、计算机、基因……）的集合，它们通过无序的多种相互作用模式而彼此关联，即它们都有着某种潜在的网络结构。通常，这种隐含的网络结构是理解前述重要现象的1关键所在。其中一个很好的例子便是上世纪80年代大西洋西北海域鳕鱼种群的衰竭。当时，鳕鱼的短缺给加拿大水产业造成了巨大的经济危机。加拿大境内的利益相关者吁求更多地猎捕海豹，他们认为控制鳕鱼捕食者的数量有助于阻止鳕鱼种群的衰竭。大量的海豹在上世纪90年代遭到猎杀，但鳕鱼种群并未恢复。与此同时，生态学家们则对关联鳕鱼和海豹的不同食物链进行了研究。到上世纪90年代末期，他们已经为研究发现的关联这两个物种的大量不同食物链条绘制了一幅详尽的网络图（图1）。根据这个复杂的网络图，捕杀鳕鱼捕食者并不必然有助于鳕鱼种群的恢复。比如，海豹捕食大约150种猎物，其中也包括鳕鱼的其他捕食者。因此，海豹种群数量的减少便增加了其他鳕鱼捕食者对其种群数量所造成的压力。

生态系统乃复杂的物种网络：如果我们想要理解并管理这些物种，将这种底层的网络结构纳入考虑范围则至关重要。我们对其他一些基于网络结构的系统也必须采取类似的谨慎态度。例如，像艾滋病之类的传染病的蔓延就受到某些人群内部无保护性关系模式的强烈影响。同样，流动性冲击则取决于银行间货币交换网络的相互交织。

上述所有例子都是所谓涌现现象的例子。此种现象乃无法通过观察构成系统的单个元素而进行预测的集体行为。通常，呈现这些现象的系统被称为复杂系统。比如，单个的蚂蚁是相对笨拙的昆虫，但众多蚂蚁一起则能够进行如修建蚁冢或者储存大量食物这样复杂的活动。而在人类社会中，社会秩序产生于自主个体的结合，尽管他们的利益往往相互冲突，但最终仍然合力执行凭一己之力无法完成的任务。类似地，某个鲜活有机体的生命力来自其不同组分的相互作用；互联网应对错误、攻击和信息量高峰的非凡弹性是网络作为整体的性能，而非单个机器运行的结果。

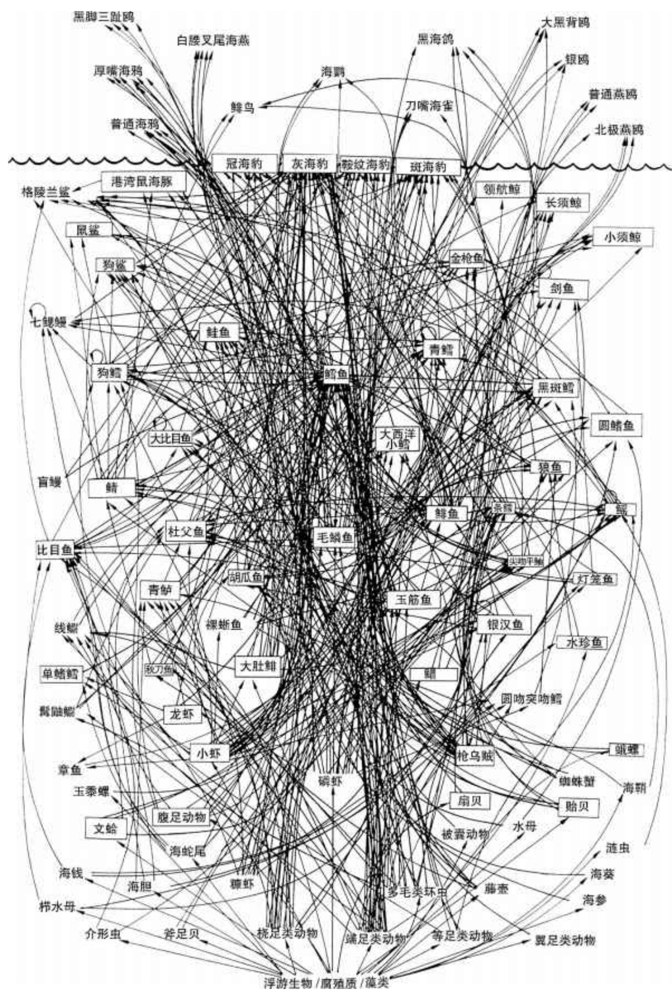


图1 加拿大东部西北大西洋“斯科舍大陆架”食物网局部图。箭头方向从捕食者指向猎物

网络及其所强调的互动，乃理解诸多此类现象的关键。设想两支足球队的球员水准十分相近，最终表现却差异甚远。这种差异很可能取决于球员们在场上相互配合的好坏程度。类似地，单个球员可能在联盟球队中表现尚好，而在国家队中却表现欠佳，因为该球员在两个团队中的相对位置有所不同。团队的表现是某种涌现现象，它不仅取决于单个参与者的才能或个体技艺的总和，而且还取决于参与者之间所形成的互动网络。许多涌现现象都严重依赖于其底层的网络结构。

网络方法将所有注意力都集中在了单个系统内相互作用的全局结构之上，而每个组成部分的具体属性均被忽略。结果，像计算机网络、生态系统或者社会群体等十分不同的系统都用同一种工具来描绘，那就是图（graph），即由多种关联所结成的一种单一的节点架构。这个方法最初由莱昂哈德·欧拉在数学领域发展起来，之后扩展至广泛的学科领域，包括大量使用这种方法的社会学，以及最近的物理学、工程学、计算机科学、生物学和其他诸多学科领域。

使用同一种工具来表现差异显著的不同系统只能以某种高度抽象的方式完成。系统细节描述的缺失在普适形式中得到弥补，即将十分不同的系统视作相同理论结构的不同实现形式而加以考虑。由此观之，计算机病毒的传播可能与流感类似；劫持一个路由器可能与某个物种在生态系统中的灭绝有着相同的影响；而万维网的增长则能与科学文献的增长放在一起比较。

这条推理线索提供了诸多洞见。比如，把系统呈现为图能让我们察觉到那些将明显不相关元素囊括在内的大规模结构。2003年，瑞士电网中的一个小事故触发了一场大规模停电事件，其影响波及1 000公里之外的西西里岛。专注网络结构让我们看到，遥远事物之间最终可能通过难以置信的关系或传播短路径而强烈地关联着。地理和社会上相互隔离的两个个体——比如某个热带雨林中的居民和伦敦金融城中的某位经理——仅通过六度分隔（six degrees of separation）就能相互关联，这个最近的观察与现实相去不远。并且，这种现象可用社会关系的网络结构加以解释。

网络方法还揭示出系统的另一个重要特征，即某些不受外部控制而发展起来的系统仍然能够随机发展出某种内部秩序。细胞或生态系统并非“设计”而成，但仍以某种牢固的方式起作用。类似地，社会群体及社会趋势源于大量不同的压力和动机，但仍然展现出清晰和确定的图景。而互联网和万维网则在缺乏任何监管机构的情况下兴盛起来，并由大量不相关的因素所推动，但它们通常还是按照一致且有效的方式运转。所有这些现象都是自组织过程，即系统内部秩序和组织并非外部干预或整体设计的结果，而是局部机制或倾向在成千上万次

相互作用中迭代产生的现象。网络模型能够以清楚和自然的方式描述自组织在诸多系统中是如何产生的。同样，网络让我们能够更好地理解诸如计算机病毒的快速传播、大规模流行病、基础设施的突然崩溃，以及社交恐惧症或音乐流行之爆发等各种动态过程。

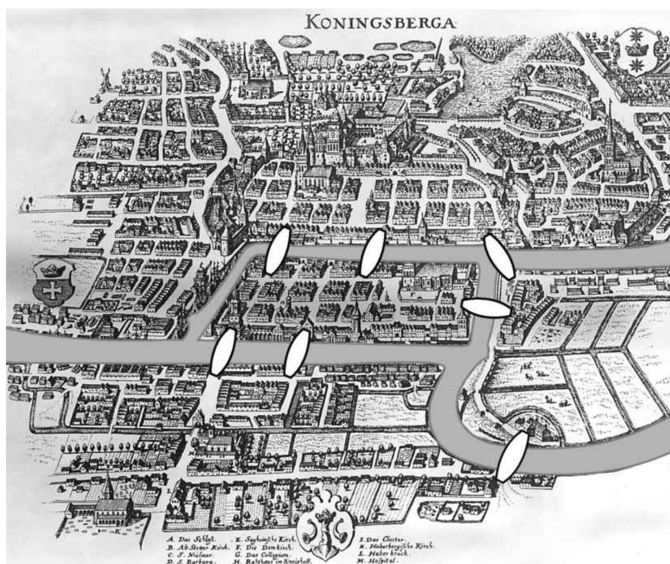
在复杂、涌现以及自组织系统的研究领域（即关于复杂性的现代科学）中，网络作为通用的数学框架而变得愈发重要，涉及大量数据时尤其如此。这通常体现于个人在搜索引擎中的累积查询、社交网站的更新、在线支付、信用卡数据、金融交易、移动电话中的GPS（全球定位系统）位置等情况中。在所有这些情形下，网络对于整理和组织这些数据，进而将个人、产品和新闻等相互关联而言是重要的工具。类似地，分子生物学则越来越依赖计算策略以便在其自身产生的大量数据中找到秩序。科学、技术、健康、环境和社会等许多其他领域也是如此。在所有6这些情况中，网络正成为揭示复杂性之隐藏结构的典范。

第二章 富有成效的方法

穿越欧拉之桥

在俄罗斯的加里宁格勒市，一座名为奈佛夫的岛屿坐落于普雷格尔河内。该市300年前曾隶属于普鲁士王国，彼时唤作哥尼斯堡，而奈佛夫岛与该市其余部分则由七座桥相互连接（图2上）。当时的城里流行着一个谜题：是否可能在不重复走过一座桥的情况下，遍历所有七座桥？之前没有任何人成功做到这一点。另一方面，对这种走法的可行性，当时也不存在任何形式证明。而其解决方案来自古往今来最著名的数学家之一。1736年，莱昂哈德·欧拉以一种不寻常的方式画出了哥尼斯堡市的地图。他将陆地和岛屿的部分表示为点，而桥则为连接各点的线（图2下）。

当我们以这种方式看待该问题，事情就变得更加容易了。通过展示城市的网络结构，欧拉证明了谜题中的走法不可能。他的解释基于以下观察：若此种走法可行，则路程沿线所有点的连接数必为偶数。这是因为每当某人通过一座桥进入该市的任何部分时，他必须通过另外一座不同的桥离开。总之，每个区域必须有偶数座桥，比如2座、4座、6座等等。只有起点和终点可以有奇数个连接数：起点只能有一座桥，终点亦复如此。不幸的是，哥尼斯堡地图所有顶点的连接数均为奇数。因此，人们不可能一次性遍历所有桥而不重复。



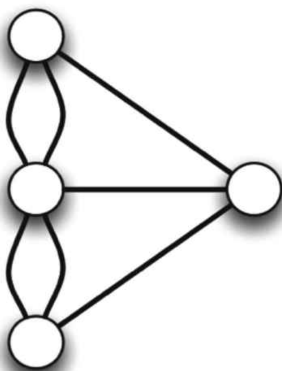


图2 版画中的哥尼斯堡（上），即现在的加里宁格勒，它表现了哥尼斯堡七桥之谜，莱昂哈德· 欧拉将其表示为一幅简图（下）

哥尼斯堡地图这种简单的数学图是图的首个例子。数学家将构成这幅图的点和线分别称为顶点和边。如今，欧拉已经因为开启了一整个建立在图解分析基础之上的数学分支而被人们铭记。他的这种直觉可被认为是网络科学的首个创立契机。在他之后，许多数学家都曾研究过网络的形式特征，科学家则将它们应用于更加广泛的问题之中：1845年G.R.基尔霍夫的电路，1857年A.凯利的有机成分异构体，1858年由W.R.哈密顿提出的“哈密顿圈”，等等。其中一个著名的应用是于19世纪中期提出的“地图着色问题”。当时，地理学家正在设法计算出绘制地图所需的最小颜色数量，其中相邻的国家应有不同的颜

色。这不仅仅是个理论问题：考虑到为数众多的国家和印刷行业中数量有限的不同油墨，使用最少数量的颜色显得至关重要。从经验角度看，三种颜色不够用，四种颜色则似乎很奏效。人们直到1976年才证实了解决方案确为四种颜色。证明基于将一张地图以图的方式呈现出来，其中的节点为国家，边则画在两个共享一个边界的国家之间。

出走的女孩，澳大利亚人和芝加哥工人

1932年的秋天，在两个星期内有14名女孩从位于纽约州的哈德森女子学校中出走。这种出走率不同寻常。学校管理层于是决定调查这些女孩的个性，以便理解这种现象。由于没有发现女孩们的性格有特殊差异的明显证据，精神科医生雅各布·莫雷诺便提出了一种完全不同的解释。他指出这种大规模的出走情形是由出走女孩们在社交网络中的位置引发的。莫雷诺和海伦·詹宁斯一起利用社会计量学，即一种能够识别个体之间关系的技术，绘制了学生之间的社会关系图。他们发现，这些关系是出走的想法在女孩们之间传播的主要渠道。个人在友谊网络中的位置对其模仿出走女孩的行为至关重要。

莫雷诺是第一批将网络理念应用于社会的研究者之一。继欧拉的直觉之后，他的研究成为网络科学创立的第二个关键步骤，它开启了网络科学最重要的分支之一：社交网络分析。30年之后，人类学家采用了类似的方法研究部落的亲属关系，比如对澳大利亚阿伦达人部落的研究。在这个案例中，互为亲属的人们之间的联系图得到绘制。研究人员发现，他们得到的结果图对应着优雅的数学结构。这些以及其他结果都表明，在人类社会的无序之下，暗含着精巧的社会结构，甚至是普遍法则。从那时起，社会科学便广泛运用网络概念来呈现社会结构。许多其他的研究也遵循着这些最初的研究方法，比如，对南美洲妇女群体（A.戴维斯、 B.B.加德纳以及M.R.加德纳等人，1941）、芝加哥工厂的工人群体（E.梅奥，1939）、学童之间的友谊（A.拉波波特，1961），以及对科罗拉多州斯普林斯市吸毒者之间关系的研究（理查德·B.罗滕伯格等人，1995），诸如此类。

随机联系

网络科学创立的第三个重要时刻则随着保罗·埃尔德什和阿尔弗雷德·雷尼这两位数学家在1959年到1961年间一系列论文的发表而到来。埃尔德什乃20世纪最重要的数学家之一，人们称他为“数字情种”。这句话不全对——他也钟爱着图形。这两位理论家研究了某种表示图的数学模型，其中的顶点完全随机地连接。数学研究人员雷·萨

洛莫诺夫和阿纳托尔·拉波波特最早于1951年在一篇论文中提出了这种后来被称作随机图的模型。

随机图是一种十分简化的模型，其特征与那些真实网络截然不同。比如，随机性和概率的确可能在结交新朋友的过程中发挥重要作用，但是友谊网络的形成肯定也与许多其他因素相关，比如社会阶层、共同语言、吸引力等等。然而，随机图模型仍然非常重要，因为它量化了完全随机网络的性质。随机图能用作任何真实网络的基准或无效状态。这意味着，我们能将随机图与真实世界的网络进行对比，进而了解概率在多大程度上塑造了后者，以及其他标准在何种程度上起了作用。

以下是构建随机图的最简单方法。我们取所有可能的顶点对，然后为每一对顶点抛掷硬币：如果结果是正面朝上，则画出一个连线；否则我们转向下一对顶点，直到遍历所有顶点对（这意味着我们以 $p=1/2$的概率绘制顶点连线，但可能会用到 p 的任意值）。通常，图的创建及相关研究都不是人们手动完成的。科学家运用计算机程序，在图纸或电脑屏幕上绘制生成网络图。然而，当网络规模很大时，这项工作就会变得越发困难。此外，人们仅通过目测很难研究相互交织的结构。运用数学工具将图作为抽象对象进行研究能够提供更好的量化的见解。计算机对此也有帮助：模拟使我们能够在计算机的“头脑”中将这种模型可靠地具象化，然后对其进行测量，就像它是个真实的物体一样。如果我们想要将抽象模型与真实世界网络进行比较，现在只需要比较两种情况下的测量数据。

自上世纪60年代被引进以来，随机图模型已成为最成功的数学模型之一，尽管它与现实的联系并不紧密。现在，它是所有网络进行比对的基准，因为任何对此模型的偏离都表明了在许多现实世界网络中存在着某种结构、秩序、规律性和非随机性。

捕获信息之网

当纽约、马德里、伦敦分别于2001、2004、2005年遭遇死灰复燃般的恐怖袭击之后，一些政府机构提议存储电子信息流数据，并将其作为一项反恐措施。在这项提议中，公民之间多年的电话和电邮都会以安全的名义被记录在案。具体的通信内容则不会被储存，仅记录信息的发出者和接收者便已足够（有时候还包括通信的时间和地点信息）。正如警方所知晓的那样，即便这种简单的谁与谁之间相互关联的图景也是追踪人们活动的强大工具。的确，略略看一下某人的电话记录，我们就可以推断出他的习惯、朋友圈以及各种相关的其他数

据。

这是网络科学基本方法的一个非常实际也颇具争议的例子。复杂系统由图——通过同样的相互作用而彼此关联的一组等价元素——表示，却忽略了其组成部分的详细特征，以及它们之间关系的具体性质。这一做法似乎过于激进，但它仍然让我们能够捕获到比乍看之下可预期到的更多的信息。这种方法的一个有效证明便是内嵌于诸多在线社交网络中的朋友推荐系统，比如“脸书”或“领英”。其中的想法很简单：你很可能认识你朋友的朋友。虽然很简单，却在多数情况下都有效。这种方法被称为有根据的推测，它也是很多在线商店的书籍或其他商品推荐系统的基础。这些公司的软件利用了与每个消费者相关的商品网络。这就是为何商业公司会储存大量的电子数据，包括电子邮件和在线社交网络数据——它们知道这种信息十分有用。

网络科学的基本方法可应用于广泛的系统之中。例如，一个由数百个物种组成的生态系统，其中每个物种通过不同的捕食策略从其他物种中提取能量。在网络方法中，这种情况由一组相同的边连接起来的一组相同顶点表示。同样的策略被应用于从互联网（通过电话线、光纤电缆、卫星通信等方式连接起来的无数计算机、路由器和交互站点等）到人口族群（通过各种关系相互联系且拥有不同目的和身份的大量行动者）的各种系统之中。尽管网络方法消除了所考虑现象的许多个别特征，它仍然保留了个体的某些特征。也就是说，这种方法并不改变系统的大小——系统元素的数量，也不会改变相互作用的模式——元素之间的特定连接集合。然而这种简化的模型仍然足以捕获系统的特性。

从个体到群体

在处理多种不同元素以不同方式相互作用的情况时，存在两种可能的方法。第一种方法是确认这种情况中的基本成分以及它们之间的相互作用。通过研究每个元素本身，我们可以将系统的行为推导为个体元素的加总。例如，生态学家可以通过列出每个物种的猎物和捕食者来描述生态系统的特征。计算机科学家通过关注每个不同机器的特征和协议来描述计算机网络。心理学家则通过描述每个社交者在他的圈子中的行为来研究社会关系。与第一种策略不同，第二种策略旨在将许多元素组合成为若干同类的群体。例如，社会学家和政治学家通常按照社会阶层、性别、教育水平、种族、国家等因素对社会进行划分。

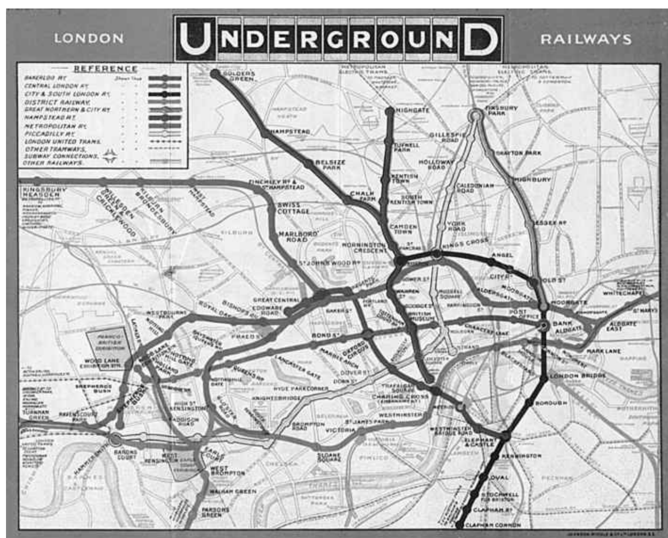
同样，流行病学家则常常将人口分成一组有限的“区划”：健康

的、感染的、免疫的个体等。生态学家也能用同样的方法将所有在食物网中扮演类似角色的物种汇集成群组（即营养物种）。

网络方法试图补充这两种策略。若我们只关注个体元素的行为，许多现象就无法得到解释。例如，生态系统内物种的数量动态并不单单依赖于该物种自身的特征，猎物–捕食者相互作用的全局网络也必须考虑在内。另一方面，专注于大类的元素也可能没什么用。比如，一国发生的政治现象基本上不是由先前存在的国家认同引起的，而是该国内部社会关系的具体模式运作的结果。网络方法则介于通过单个元素和通过大群组进行的描述之间，并将二者连接起来。在某种意义上，网络试图解释一组孤立的元素是如何通过相互作用的模式转化为群体和共同体的。在与这种模式相关的所有情况下，网络方法都提供了重要的洞见。

地理与“网络图”

20世纪初，伦敦的地铁服务变得十分复杂，以至于人们必须不时地发布愈来愈多的地图以便为出行者提供向导。1931年，经过多次尝试，地铁公司的雇员亨利·贝克改变了绘制地铁路线图的标准。与那种将地铁路线嵌入伦敦实际地图之上的做法相反，贝克将它们置于抽象的空间中（图3）。站点由间隔合适的点表示，而线路的连接则全部变成了具有45或90度角的直线。这个地图无关车站的真实位置及其相互之间的实际距离，但它对乘客而言更加清楚且有用。乘用地铁网络的人对其地理特征不感兴趣，有车站的顺序和地铁路线的交汇信息便已足够。



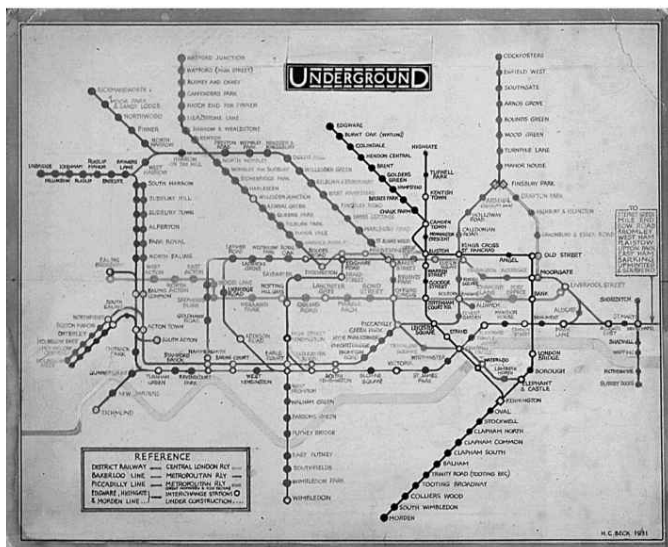


图3 伦敦地铁的“度量”示意图（上）与“拓扑”示意图（下）。尽管车站的实际位置和相对距离都未呈现在拓扑图中，它却是地铁服务的更佳思维导图

亨利·贝克的伦敦地铁图基本上就是一个图。他解决映射问题的方案利用了网络方法的一个基本特征：在网络中，拓扑比度量更为重要。也就是说，何物与何物相互关联比它们相隔多远更加重要。换句话说，图的实际地理情形不如其“网络图”重要。这两个概念的差异如图4所示。从度量角度看图中呈现的三个图形有所不同。也就是说，

空间中的节点位置和节点连线的长度有所不同。然而，从拓扑学的角度看，它们是相同的——它们只是同一个图的三种不同呈现方式。在网络的表示方式中，系统元素之间的关联比它们的相对距离以及在空间中的具体位置要重要得多。

对拓扑学的关注是网络方法的最大优势之一，当拓扑与度量相比更为适用时，网络方法就是有用的。例如，从纽约发送至伦敦某办公室的电子邮件会与来自其隔壁办公室的邮件同时抵达。即便在互联网这种嵌入地理空间的实体基础设施中，连接的模式也比物理距离更加重要。在社交网络中，拓扑的实用性意味着社会结构很重要。里奥内尔·梅西是当今世界最好的足球运动员之一。然而，他的表现却随着他所效力球队（比如阿根廷国家队或者巴塞罗那足球俱乐部队）的不同而有所不同。一些社会科学家认为，梅西在阿根廷国家队中与其他队员之间的关系网络不同于他在巴塞罗那足球俱乐部队中的关系网络。根据他们的研究，这会导致球员在前一种情况下背负更重的“负担”，这至少能部分解释他在不同球队中的表现差异。类似的现象也出现在更加复杂的社交“游戏”中，个人在其中的境况会受到他或她在关系网络中所处位置的强烈影响。

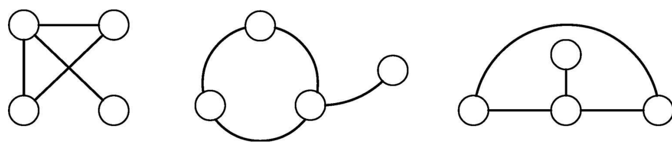


图4 同一个图的三种不同表示

链条、网格和网络

网络方法将复杂系统简化为单一的点线架构。这是一个重大的简化，但其得出的结果图却可能不是那么容易得到解释：这便是图4所示的棘手例子展示的情况。甚至一幅仅由节点连接而成的链条图也可以是一个处理起来相当复杂的对象。一个链条可以代表比如搬运一桶水的消防队；或者物种的食物链，其中，第一个捕食第二个，第二个捕食第三个，以此类推；或者企业之间的供应结构——一系列公司中每个公司供给下一个公司。

想象一下五家公司（公司1，2，3，4和5）的生产链。在这个链条沿线，其中任何一家公司都能与其两个邻居之一达成交易。其中的规则为，每个公司只能签订一份合同——例如，如果公司3与公司2达成了交易，则它不能与公司4另有协议安排。给定了这个简单的结构和规则，结果便是位于节点1和节点5的公司议价能力更低，因为它们

的选择更少。这导致公司2和公司4更加强大，并且（意外地）削弱了位于节点3上的公司。确实，节点3上的公司只得应付更加强大的公司，因此，它最终只能达成不那么有利的交易。简单如节点线性序列这样的事物确实产生了相当复杂的景观。社会学家将这个例子呈现的现象称为排斥机制。这远非理论中的情况，而是经济学中经常能遇到的情形——当两个公司之间建立商业关系以排除第三方时便是这样。

为了进一步深入这个问题，人们必须考虑到现实中的系统绝少如链条这般简单。在过去，科学家用规则网格或晶格表示复杂的系统，而非使用图。这些对象由许多元素组成——分别代表人、动物、计算机等等——它们通常按照规则的连接模式排列，例如棋盘上的棋子仅与其四邻的棋子相连。相对于用图表示而言，这种规则结构使得系统能够更容易被数学计算和计算机模拟这两种方式处理。

尽管格子相对于图而言更加简单，但它引入了强烈的限制条件。事实上，一个格子仅适于呈现经过精心设计或受到强烈约束的系统。比如，这些系统可能是某个计算集群的处理器阵列，或者某个嘈杂工作场所中彼此关联的工人之间的语言交流。

在格子框架中，每个节点都连接着固定数量的最近连接点，而在绝大多数真实情况中，这些连接则关联着数量可变的元素，无论它们彼此是邻近还是相隔很远。捕获这种混乱状态的能力是网络方法的最大优点之一。

大部分这种无序状态都被编码为某种关键的数量——度数，即附着于每个节点之上的边数。如果节点是网页，度数则表示了该网页与其他页面的链接数量。如果节点是物种，度数则表示了该物种赖以生存的物种数量。如果节点是一个人，度数则表示了其熟人的数量。这个圈子能与社会学家彼得·马斯登所谓的核心讨论网络联系起来，即那些可与之讨论重要事宜或打发时间的人群（朋友、伙伴、家庭成员、现在和过去的同学、同事、邻居、俱乐部伙伴、专业顾问、咨询员等等）。

描绘关系

两个人之间可以有无限的可能关系。他们可能共有同样的态度、想法或性别，可能是朋友、亲戚或同事，也可能是性伴侣或者仅仅效力于同一个足球队。此外，这些关系中的两个或更多可以同时发生在同样两个人身上。其中一些是合作关系，另一些则表示公开的敌对，而介于两者之间的关系则有着广泛的可能。最后，一些关系可能仅仅

单方面被感知到，而另一方则完全对其忽略（比如，摇滚明星的粉丝可能感觉自己与其有某种联系，而摇滚明星可能完全忽略了他们）。社会学将个体之间广泛的可能联系进行了分类（表1）。在社交网络中具有多种关系类型的倾向又被称为关系的多重性。但这种现象也出现在许多其他的网络中——例如，两个物种可以通过不同的捕食策略而彼此关联，两个计算机则通过不同的有线或无线连接彼此关联等。

我们可以修改某种基本图来考虑关系的多重性，例如，为关系连线贴上特定的标签等。比如说，我们可以将某种关联是正面还是负面纳入考虑范围。物种通过捕食（负面）或共生（比如在开花植物和授粉者之间建立起的正面关系）而彼此关联。人们可以是敌人（负面）或朋友（正面）。一个网页则可以连接到另外一个网页以批评其内容（负面）或为其做广告（正面）。通过增添这个简单的二进制功能便让事情极大地复杂化了。想象一个由爱丽丝、鲍勃和卡罗尔组成的三人群体。当他们所有人由正面的关系联系时，一切都安好。又或者，爱丽丝和鲍勃可能是朋友关系，但他们二人均对卡罗尔抱有敌意。当情况相反的时候，事情就变得复杂了：爱丽丝与鲍勃和卡罗尔二人都保持着正面的关系，但鲍勃与卡罗尔之间却彼此憎恶。而当他们每个人都彼此憎恶的时候，情况就变得着实棘手了。根据社会学理论，第一种情况和第二种情况在结构上是平衡的，而第三种和第四种情况则相反。2006年，数学家蒂博尔·安塔尔和物理学家保罗·克拉皮夫斯基、西德尼·雷德纳将这个概念应用于第一次世界大战前六个欧洲国家之间外交联盟关系转变的分析之中。他们证明了这六国的联盟逐渐演化成了一种结构上平衡的状态，在这种情况下，强大的联盟得以建立，或者明确的共同敌人得以确定。这六国后来分裂为两个集团（一方为英国、法国和俄罗斯；另一方则为奥匈帝国、德国和意大利），每一方的成员国彼此在集团内相互结盟，而对另一方的所有国家保持敌对关系。这种情况出现后不久，战争便爆发了。这个例子表明，结构上的平衡并不必然就是有利的。

相似点			社交关系				互动	流动物
位置	成员	属性	亲属关系	其他角色	情感关系	认知关系		
例如： 处于同一个时间和空间	例如： 同属一个俱乐部 经历同样的事情	例如： 同样的性别 同样的观点	例如： 母亲 兄弟 姐妹	例如： 朋友 老板 学生 竞争者	例如： 喜欢 憎恶	例如： 了解 知道 认为 …… 幸福	例如： 与某人有性关系 与某人交谈过 给某人建议 帮助某人 伤害某人	例如： 信息 信念 人事 资源

表1 社交网络关系的分类

图论让我们能为边编码更加复杂的关系，当关系并非对称的时候更是如此。狼捕食羊，博客链接到大报，一些人爱上另一些人；相反的情况绝少为真。在这种情况下，图中的关系乃某种单行道，我们可在其中朝着一个方向前进，但不能逆行。如果给边添加方向属性，则得出的结构为有向图，其中的关系由箭头指示。在这些网络中，我们有入度和出度分别测量节点向内与向外的连接数量。

到目前为止，我们所考虑的关系都是二进制的——它们仅有两个值。这种二分关系要么存在，要么不存在：例如，与某人结婚或者被某人雇用。然而，这种关系在统计上都是例外：绝大21多数情况下，关系展现出广泛的强度变化。捕食以吃掉猎物的数量计算，网页则能被零星或大量的链接关联，而爱情的强度则从轻微的吸引到强烈的激情不等。这些进一步的特征对应于我们赋予关系的权重。例如，加权网络可能作为个体或事物之间不同交互频率的结果出现。

我们对基本图结构的其他修改也是可能的，且处理这些事物的技术非常有趣。比如，大部分社交网络研究致力于了解不同种类的关系如何互相影响。然而，网络方法的优点在于，在某些情况下，忽略所有或大多数具体细节是合理且有效的：有向网络变得无向，权重被移除，多重连接坍塌为某条单独的边等等。结果表明，这种彻底的简化仍然能够捕获大量信息。

第三章 网络世界

网络组学

上世纪80和90年代，人们倾向于认为在某种方式上一切都是由基因决定的。报纸上的报道都与“同性恋基因”“肥胖基因”“暴力基因”或者“酗酒基因”有关。这种态度呼应了人们对人类复杂性的秘密隐藏在基因组中的期望。DNA——细胞核中包含基因的脱氧核糖核酸分子——又称“生命的软件”，该程序负责生命体的每一个特征正是其中代码功能的障碍导致了所有的疾病。这种图景引发了人们对基因组进行测序的风潮，而人类基因组图谱于2001年2月的发布将这个风潮推向了顶点。测序的结果令人十分惊讶，人类的基因数并不比线虫多多少，而且比某些种类的水稻基因还少。人类几乎与类人猿有着相同基因组的推断是合理的，但问题是人类基因组也与老鼠类似。软件的隐喻并不支持这一证据：DNA序列本身并不能解释我们所能观察到的物种差异，更不用说单个个体的所有特征和疾病了。事实上，生物的基因与其相应的宏观特征之间还隔着一系列漫长的步骤。此间的变异决定了不同的结果。

基因之上的第一层复杂性由基因调控给出。包含在DNA中的基因被转录和翻译以产生蛋白质。蛋白质几乎在生命的各个方面都起着核心作用：肌肉运动、血液循环、承担酶的作用、结合激素等等。此外，蛋白质还彼此相互作用：蛋白质的产生可能受到细胞内现有蛋白质的促进或阻碍。这些相互影响之间的微妙平衡对于生命而言至关重要。例如，p53这一单个蛋白质的突变就伴随着大量不同的癌症。这些交织的激活和抑制模式产生出了基因调控网络。在这个网络之中，节点为基因，连线则是关联某个基因与其他基因表达的反应链条。

蛋白质以所有可能发生的方式相互作用，这代表着另外一个层面的复杂性。例如，一些蛋白质可以结合在一起。这些大分子像分子机器一样起作用，它们在细胞这种机器中发挥着自己的功能。为此，蛋白质必须具备正确的几何形状以彼此契合。当蛋白质以错误的方式折叠时，就可能出现一些问题。比如，人体中对应“疯牛综合征”（又称克雅二氏病）的蛋白质（即朊病毒）就被认为是错误折叠的。蛋白质之间所有可能的实体联系可表示为网络。在蛋白质相互作用网络中，

顶点为蛋白质，而如果蛋白质在细胞中的确相互作用，则可在相应的点之间画出一条边。

蛋白质并不足以让细胞工作。细胞通过许多不同的分子与外界环境交换着物质、能量和信息，参与其中的反应数达百万级。饥饿、饱腹感、寒冷以及生物体大致上所经历的所有状态都赖于这种被称为新陈代谢反应的集合。而通过一系列中间步骤将一个分子转化为另一个分子的反应链则被称为代谢途径。然而，细胞中的反应很少遵循有序序列的模式。例如，反应终端的分子经常会与开端处的分子相互作用以终止反应。这种反馈过程就闭合了反应链中的一环。所有这般路径的集合就产生了复杂的代谢网络。

因此，生物体是几层网络共同作用的结果，而不仅仅是简单基因序列的决定性结果。基因组学中已经加入了表观基因组学、转录组学、蛋白质组学、代谢组学等研究相关层级的学科，这通常被称作组学革命。而网络正是这场革命的核心。

思维之网

直到18世纪，“灵魂”可以体现在某个器官中都是一个奇怪的设想。但是，医生已经意识到中风或者其他脑损伤可能会危及关键的认知功能——心与脑之间的关联从那时开始变得明显起来。当时，解剖学家弗朗茨·约瑟夫·加尔便敢于提出所有心理官能必然起源于大脑这种想法。他确定了大脑中的27个“器官”，每个器官分别负责颜色、声音、记忆、言语，以及友谊、仁慈、骄傲等等。这种想法听起来如此异类，以至于加尔不得不逃离维也纳，进而在激进的法国找寻避难所。

后来，几位生理学家试图通过比如从鸽子的大脑中去除切片这样的方式验证加尔的理论。然而，他们并未找到加尔所设想的器官存在的任何证据。因此，他们得出结论说，大脑是产生思想的一个均匀、未经分化的统一体：如其中一位生理学家所说，“大脑分泌思想就像肝脏分泌胆汁一样”。这种观念在1860年代保罗·布罗卡的研究出来之前一直占据主导地位。在表达性失语症患者的尸体解剖中，布罗卡总能在他们大脑左侧的额叶区域发现一些损伤。在确定了人们现在称其为布罗卡区域的大脑部分之后，他便宣称，“我们用左脑说话”。从那时起，神经学家找到了负责不同活动的多个功能中心，但他们也发现这些中心区域绝少孤立运转：大脑不同区域的整合对其功能而言至关重要。

网络在被划分为专门领域的大脑模式和作为整体的大脑模式之间架起了桥梁（这与社会科学的情况并无不同，在这个领域中，网络允许我们从个人和共同体之间的某个层面描述社会）。大脑中遍布网络，其中的各种网状结构在专门的区域之间提供了整合。在小脑中，神经元形成了一个个不断重复的模块：它们的相互作用被限制在了相邻的模块之间，这与晶格中发生的事情类似。而在大脑的其他区域，我们发现了随机连接现象，即局部、居间或者遥远神经元之间的连接概率大致相等。最后，新皮质——哺乳动物的许多高级功能所涉及的区域——则将局部结构与更加随机且遥远的连接相结合。一些科学家认为这些线路体系可能负责主体意识：涌现的良知便可能是某种足够复杂的网络结构作用的结果。

确定这些神经网络的实际结构十分困难，因为细胞数量巨大，并且探测它们困难重重。我们仅掌握了那些十分简单的生物的详细神经网络图，比如一种名为秀丽隐杆线虫的寄生线虫。这种一毫米长的透明生物能存活三周，它仅有300个神经元，却是分子生物学界的超级明星。秀丽隐杆线虫是一种模式生物，即特别适于实验的生物，因为科学家对其特征了如指掌，且它在某些方面与人体类似。这种半透明的蠕虫通常是新药和新疗法实验的第一个基准。

目前，人们还不可能为人类大脑绘制类似的神经网络。然而，我们可以使用另外一种策略。当人们执行一个动作，即便简单如眨眼，来自神经元的电信号风暴就会在大脑中的数个区域爆发。这些区域可以通过诸如功能性磁共振这样的技术加以确认。通过这种技术，科学家发现不同的区域会发出相互关联的信号。也就是说，它们显示出某种特殊的同步现象，这表明它们可能相互影响。可将这些区域视为节点，而如果它们之间存在足够的相关性，则它们两两之间便可画上一条边。同样在这个层次上，大脑表现为相互连接的元素集合。人的每个动作都会点亮脑中的某个连接区域。

盖亚的血管

1999年，旧金山湾区经历了大规模的藻类暴发。通常，藻类的这种暴发是农业密集用地的结果：我们将氮、磷等化肥排入海中后，它们就会成为藻类的营养物。然而，本案例中的情况并非如此，因为一些政策和限制已经减少了从不同河流排入海中的营养物污染水平。加利福尼亚州的生态学家们通过搜集三十年的观察数据得出结论说，藻类暴发有着复杂得多的原因。1997年和1998年，人们记录到了最强的“厄尔尼诺”现象之一，紧随其后的便是1999年同样强大的“拉尼娜”现象。这些现象导致了当时加利福尼亚生态系统的变化。深海处

冰冷的富营养化海水涌至海岸。这些营养物质吸引了海洋中的生物——鲽鱼和甲壳类动物——进入湾区。这些动物是湾区双壳类动物的捕食者，而后者反而是藻类蔓延的障碍。双壳类动物由于捕食者的27增加而种群崩溃，这成为藻类暴发的直接原因。触发这种多米诺效应的条件可能来自正常的气候波动。然而，其后果对人类而言却是一种警告：气候变化，特别是极端气候的频率增加，可能对生态系统产生意想不到的影响。

旧金山湾区海藻暴发背后的核心结构是食物链——一系列物种的连接：鲽鱼和甲壳类动物捕食双壳类动物，双壳类动物消耗藻类。生物以食物链的方式从彼此身上提取他们生存所需的能量和物质（这并非生态系统内物种之间唯一可能的相互作用：生物体也可以建立互利的相互作用，比如开花植物与其授粉昆虫之间的相互作用）。每个食物链都从基位物种开始，比如植物和细菌。这些物种不会捕食其他物种，它们通过转化阳光、矿物质和水而直接从环境中获取资源。这些资源以接连的捕食方式沿着食物链顺次转移。中位物种既是捕食者也是猎物。而顶位物种（处于食物链顶端）则不被任何别的物种捕食。食物链帮助我们理解为何渔业会崩溃，比如70年代秘鲁鲥鱼产业的境况。在大规模的无差别捕鱼期之后，像鳕鱼或金枪鱼等捕食者就会大幅减少。在这个阶段之后，捕鱼业则朝着更加基位的物种进发，比如鲥鱼。但这些物种也会迅速崩溃。原因在于，当大型捕食者被移出食物链之后，食物网下游的其他捕食者便会取而代之。后者往往是不可食用的鱼类：它们的种群没了限制之后，便会耗尽其他可食用的基位物种。

生态系统的实际情况甚至更为复杂：通常，食物链并不孤立存在，而是以复杂的模式彼此交织，其中的某个物种会同时属于多个食物链。例如，一个特化物种可能只捕食一种猎物（或者在一些情况下仅捕食少数物种）。如果猎物灭绝，这个特定物种的种群就会崩溃，并导致一系列的共同灭绝。更加复杂的情况是，杂食动物捕食某种食草动物，而两者都吃某种植物。杂食动物种群的减少并不意味着植物的繁茂，因为食草动物将从这种物种减少中受益，进而消耗更多的植物。

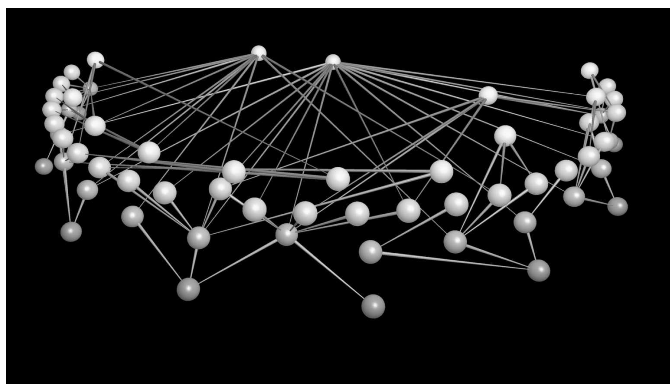


图5 英国某草原食物网：节点代表英格兰和威尔士一些草地中的物种，连线则由捕食者（较粗的一端）指向猎物（较细的一端）

将更多的物种纳入考虑范围，种群动力机制会变得越发复杂。这便是为何对生态系统而言，食物网（图5）是比“食物链”更为恰当的描述。这些网络中的节点是物种，连线表示它们之间的捕食关系。连线通常是有向的（大鱼吃小鱼，而非相反）。这些网络提供了物种之间的食物、能量和物质交换，并因此构成了生物圈的循环系统。它们构成了地球的血管。

人类“角斗士”（网络人）

口口相传（word of mouth）是获得有关白领职位空缺信息的惯常方式。所以，如果我们正在寻找这种工作，在亲友之间传递相关信息则不啻为一个好主意。不那么明显的是，我们将信息告知远方不常见面的熟人甚至会有更好的效果。这便是马克·格兰诺维特于1973年提出的观点。这位社会学家采访了波士顿郊区职业人士这一样本，他们当时都依靠个人联系获得了各自的工作。格兰诺维特问他们在获得工作之前与提供工作的人的联系频率。多数人的答复是“偶尔”，许多人的回答是“绝少”。与亲密的朋友相比，工作机会更可能来自旧时的校友、过去的同事和以前的雇主。运气或共同的朋友则是再现这些联系的管道。格兰诺维特将这种现象描述为弱连带优势。

格兰诺维特通过描述一位假想中的名为埃戈的人的熟人圈来解释这种结果。埃戈每天都和家人及一些亲密的朋友生活在一起。所有这些人很可能也彼此联系紧密。其结果便是，信息会在这群人中快速传播。因此，埃戈可能知道这群人中的所有新鲜事。与此相反，较弱的人际联系也会将他与遥远的人联系起来。这些人并不会受到埃戈社交环境的限制。因此，他们向他开放了一组新的群体集合，每一群人都

封装着别的群体无法获取的信息。弱联系的缺失会为组织、公司或机构带来各种困难。信息和技术被限制在一个群体之内，而没有抵达那些需要它们的人手中。因此，这些东西需被重新创造或从外部顾问处有偿获得。据报道，惠普的前首席执行官曾感叹道：“惠普不能闭目塞听！”格兰诺维特的直觉后来发展成为社会资本理论。

这种想法意味着一个人的联系人（以及这些联系人的联系人）使他或她能够获取那些最终可以提供更好工作和更快升迁的资源。更普遍的是，个人在他或她的社交网络中的位置对于决定其未来的机会、限制和成果而言至关重要。

衡量交情并不容易，因为它是十分主观的事情。通常，像公司流程图之类的图谱并不是很有用，因为它们并不对应着员工之间的实际关系。正因为如此，流程图并不能帮助我们了解公司内部的信息渠道（以及可能的瓶颈）。科学家们设计了大量的替代策略来绘制社交网络，从问卷调查到雪球式抽样，即系统中的每个受访者都推荐自己圈子里的某人为下一个采访对象。

这些策略使得数据的搜集就像不同个体（从美国中西部的高中生到非洲布基纳法索的村民）所组成的群体之间的性行为图那般敏感：有关这些关系网络的知识能让人们更好地理解性传播疾病的蔓延。

另外一种相对容易定义的关系则是专业协作。这样的网络存在于多个领域，从好莱坞（如果两个演员出演同一部电影，他们便会产生联系）到科学领域（如果两位科学家共同撰写一篇论文，他们也会产生联系）。合作还能在更不寻常的环境中找到，比如政治（美国参议员会在支持共同法律的基础上产生联系）或者恐怖主义活动（恐怖分子则基于情报和法律文件相互联系）中。

信息技术为衡量人际互动提供了一种强大的新手段。两人之间频繁的电话和邮件往来，或者在脸书、领英等虚拟社交网络中的友谊等，都表明了一种稳定的关系和网络图中的一条边。越来越多的公司利用他们客户的社交网络来找寻这样的信息。例如，据报道，电话公司会以提供工作机会或别的策略来瞄准“有影响力”的个人：当这些客户改换公司的时候，就会在他们的密切联系人中触发类似的变化。

语词之网，思想之网

玛格莱特王后在莎翁的《亨利六世》三部曲的中篇（第一幕，第三场）里说道：“难道说，亨利王上老要在乖戾的葛罗斯特管辖之下

当小学生吗？”[1]1王后抱怨葛罗斯特公爵对其国王丈夫的影响。她用“小学生”（pupil）表达的是什么意思？分类词典表明，“pupil”还可以表述为scholar（学者）、acolyte（助手）、adherent（信徒）、convert（皈依者）、disciple（门徒）、epigone（追随者）、liege man（臣下）、partisan（党羽）、votarist（拥护者）或者votary（崇拜者）等等。这个清单为“一个人从属于另外一个人”这项语义提供了一幅全景语词图谱。我们可以通过探索与“pupil”相关的语词即其语义区来扩展这一图谱。其他的语词包括faithful（信徒）、loyalist（忠诚者）、advocate（倡导者）、backer（赞助者）、supporter（支持者）、satellite（随从）、yes-man（唯唯诺诺之人）……那么，王后意欲用何种意义指谓其夫君呢？“pupil”的反义词表示non-student（非学生）、coryphaeus（领导者）、leader（首领）、apostate（变节者）、defector（背叛者）、renegade（叛徒）、traitor（背信弃义之人）以及turncoat（叛贼）等。当然，她在要求亨利国王反对葛罗斯特公爵的权威。

这是语词之间如何相互联系的一个简单例子。事实上，严格的分析应该将语词使用的历史差异、莎士比亚作品中某个语词的特定出现、它在剧本中使用的语境，以及其他诸多方面都考虑进来。公平地讲，在任何情况下，我们都可以更好地理解语词的含义，若我们将其在语言中的“邻居”考虑进来的话。同义、反义以及语义联系仅仅是语词间少数几种可能的联系。其他则包括整体-部分关系和上下位关系等（“啤酒”是“饮料”的一部分，后者则是前者的上位词）。“pupil”为语词间的另一种联系提供了清晰明了的例子。这个语词有着迥然不同的两种含义：它表示学生，也表示眼睛的一部分[2]。这便是一词多义的情况。自然，莎翁的戏剧语境当下就决定了其正确的含义。一般来说，上下文提供了语词的具体含义：句子中语词的共现定义了它们自身的含义。如此的共现提供了语词的另外一种关系。例如，语词“国王”和“亨利”在英语句子中就比“国王”和“相对论”一起出现的可能性要大得多。

我们现在可以创建自己的语言地图。我们用语词作为顶点，同义词、反义词和多义词（这些语词关系可按照分类词典或一般词典绘制，而共现的模式则可从大型语言数据库中获取，比如英国国家语料库）则由边来连接。语义联系更加难以确定：

对它们的研究构成了一个完整的语言学领域。一些语言具备将一个语词与一组相关语词联系起来的特殊词典。另外一个替代方案便是词汇联想实验。一个语词被提供给样本中的一群人，要求他们说出听到这个语词之后想到的第一个语词。得出的语词接着又被用来重复这

个联想实验。以这种方式，我们可以一步步建立自己的语词联系之网。语词网络的不同实例代表了不同的结果。这取决于语言本身、文本的类型、文本作者所受的教育，或许它与语言功能障碍也有关系。

语词网络包含大量信息，但通常，这些信息在研究文本的具体内容和不同文本所表达的思想之间的关系时并不是特别有用。这是一个关键问题，比如对网络查询来说，通常，人们必须引入复杂的算法来执行该任务。然而，在一些文本中，人们可以绘制十分精确的网络。科学文献中即是如此。生产知识从来不是单个人的努力。如哲学家沙特尔的伯纳德在12世纪首次指出的那样，科学家总是“站在巨人肩膀上的矮子”。科学家的工作几乎总是建立在先前的结果之上。研究者们认识到了这一点，他们会在其论文末尾引用一些较早的出版物。引用提供了对过往相关成果的认可，赋予新的成果以可信度，并参考那些在某项研究中被接纳为有效或被批判过的事实、技术和实验。近年来，出版物在很大程度上实现了标准化：文章大多以英文写成，控制方法已经同质化（主要以同行评议的方式），文章影响力的衡量标准也已制定等。同时，出版物的大型电子数据库业已建立，其中每天都会增加上千个新项目：文章、书籍、专利和工程等等。所有这些生成了一个大型的出版网络：如果一个项目引用另外一个，则它们相互关联。我们还可从这些数据库中识别作者身份，并创建科学家之间的协作网络。人们越来越多地利用这些系统绘制知识发展和科学最活跃领域的图景，并将其可视化。

互联的货币

2008年，美国的一些大型金融机构突然破产。几个月的时间里，大多数发达国家都卷入了这场史上最严重的金融危机之一。关于这场危机的原因已经有过很多分析了。可以肯定的是，它表明经济在全球层面的相互关联已十分紧密。

古典经济理论将经济参与者表示为独立、完全理性的行为者，他们专注于最大化自己的收入。然而，事实表明个人、公司、机构和国家并非相互独立：每一方都以多种方式彼此影响着。他们的行为远非完全理性，而是强烈地依赖于主观态度、情感和相互影响。

贷款是公司和机构能够紧密互联的方式之一。一个有趣的例子是私人银行间日常会进行货币交换，以便能够满足银行客户的可能要求（因此变得更具流动性）。如果客户的要求超过了某银行的流动性储备，该银行可以向其他银行借款。世界各地的中央银行会要求其他银行将其存款和债务的一部分置于它们那里，以创造一个应对流动性不

足的缓冲器。在这个意义上，中央银行确保了银行系统的稳定性，从而避免了流动性冲击。同业拆借网络的冻结是2008年金融危机的最初信号之一。

与贷款相比，持股，即某公司资本直接加入到另一个公司中，能够带来甚至更加紧密的联系。这意味着一家公司持有第二家公司的一部分，并能对其主要决策施加影响。当某公司持有大部分股票，或当它能够决定董事会多数成员的投票时，持股就转变为控股。在这种情况下，法律上独立的公司便转化为商业集团。通常，这些集团展现了某种金字塔结构，其中控股公司的权力在顶部，运作公司则处于控制层级的底部。

在大多数国家，商业集团以具体形式存在，并在法律上受到监管；但是更软性和更少规制的影响形式也可能存在。这在董事会内部最为常见。管理者经常同时在多个公司的董事会任职。显然，他们扮演着董事会之间信息、联盟或者利益的交换渠道。他们在不同董事会中的同时任职为相关公司之间建立了连锁关系。如果这些公司是明确的竞争者关系，这种情况则明显与自由市场不兼容。共同的董事要么支持其中一个公司而打压另一个，要么在公司之间建立垄断联盟（这通常会被法律排除掉）。一般而言，这样的董事会发现自己身陷来自不同公司的投资者之间的利益纠缠之中而难以运作。

公司之间相互关联的进一步证据来自股票价格的相关性。财务人员知道在同一行业（例如采矿、运输、服务、食品等）经营的公司的股票会以某种类似或“同步”的方式改变价格。例如，电子行业（或任何其他行业）内所有公司的股票价格往往同时降低或升高。金融分析师则有兴趣知晓某只股票的价格变动在何种程度上受另一只股票价格变动的影响（总之，他们想知道股票价格之间的相关性）。如果这种联系足够强，则很可能这两个公司以某种方式相互关联着。贷款、持股、共享董事或股票价格的相关性是公司之间能够建立网络的主要标准：当出现这些情况之一时，则绘出网络图中的边。

互联互通远远超出了某个特定市场中的公司。正如金融危机所展现的，事件会迅速地从国家市场范围扩展为全球局面。这种情况能够发生的一个明显途径是国家之间的进出口贸易关系。这种世界贸易网是一种以国家为节点，以贸易关系为边的网络。就像细胞一样，经济依赖于这些网络的多层结构。

关键的基础设施

2003年9月28日晚，整个意大利的灯火都熄灭了，唯一的例外是撒丁岛。恢复正常供电耗费了数小时，有些地方甚至花费了数天时间。调查发现，这场停电由发生在意大利与瑞士之间高压线路邻近处的树木闪燃引发。由此造成的供电不足导致人们对剩余线路的供电需求猛增。结果，这些线路崩溃了，并在整个电力系统中产生了某种涟漪效应。

大规模的停电揭示了电网的连通性。这些系统跨越很远的距离将电力从中心点传输至城市和工业区域。一开始的电网规划很周密，这些电网随着时间的推移而越发复杂。如今，由高压输电线路连接的发电机、变压器和变电站组成的电力网络跨越数个地区，通常是数个国家（正如2003年的例子所展现的那般）。很明显，这种网络需要仔细维护以防止危急情况。

类似的不稳定性出现在各种其他基础设施中。电话网络这样的通信系统便是一例。但可能最敏感的是交通网络：街道、高速公路和连接城市的铁路，运输燃料和其他货物的船只网络，最重要的则是机场网络。飞机每年运输数十亿乘客以及成吨的货物。这种基础设施中的一次微小故障都会产生重大后果：据欧洲空中航行安全组织估计，航班延误给欧洲国家造成的损失仅在1999年就高达2 000亿欧元。在全球化的世界中，交通网络的作用类似于生物体内的循环系统。

如世界般巨大之网

1969年10月，一条信息首次通过电话线从一台计算机传至另外一台。加利福尼亚州的两个大学实验室位于这条电话线的两端。几个字母的传输之后，信息中断了，但连接业已建立：互联网的前身阿帕网（Arpanet）诞生了。关于计算机网络的设想在过去十年里一直都有。50年代末，美国国防部高级研究计划署（US Advanced Research Project Agency，ARPA）要求工程师保罗·巴兰设计能够抵抗攻击的通信架构。尤其是即便在其部分正遭受破坏的情况下，整个系统还必须继续工作。巴兰恰当地设计了具备这种特征的分布式系统——但策略上的调整使得他的开创性研究被束之高阁。然而，在60年代，一些大学为了其他目的而要求ARPA资助一个类似的项目。学术机构渴望相互连接它们的计算机，以便汇聚其计算能力。

1969年，阿帕网连接了加州大学洛杉矶分校和斯坦福研究所。两年之后，阿帕网的节点数已超过40个，包括一些公司以及其他大学。这种通信架构是如此成功，以至于在70年代，由粒子物理学家、天文学家和一些企业创建的类似网络，如高能物理网、Span网、远程网等

纷纷出现在了世界其他地方。如果说一开始的问题在于连接计算机，后来则转向了如何连接各种网络。网际互连成为许多计算机科学家的座右铭。70年代末，工程师罗伯特·卡恩和数学家文顿·瑟夫开发了传输控制协议/互联网协议（TCP/IP）：无论何种网络内部架构，这个软件都允许它们彼此交换信息。这个代码基于开放体系结构的概念，并被应用于公共领域。最终在80年代，**TCP/IP**转换完全实现，互联网——“众网之网”——由此诞生。

这一结构很可能是最能体现网络思想的人工制品。一台连接到互联网的计算机即成为众多主机之一。如果要向特定地点发送电子邮件，我们并不需要直接与该目的地相互关联。从原点到我们的目标，信息沿着路由器这种负责传送数据包的设备传播。大量的连接使结构内部相互关联：光纤、电话线、卫星连接等。因为没人筹划过哪里应该增加主机和连接，人们并未记录互联网的大致结构。实际上，绘制主机层面的图景是不可能的，我们只能在路由器层面对其进行大致的呈现。在这种情况下，这些网络的节点便是路由器，它们的连接就是边。我们甚至可将结构更加粗粒度化，将路由器编组至自治系统中。这些组便是自主管理的域，通常对应于互联网服务提供商和其他组织。

互联网的巨大成功在于它提供的卓越体验。观看电视乃单向、单媒体的被动体验。互联网并非如此。人们可以浏览无数的文档，使用不同的媒体，交换信息，并相互交谈。与传统的通信技术，比如电话、无线电或电视不同，互联网并没有特定的目的。相反，它是一个能够容纳无数设备的异形人工制品。实际上，互联网仅仅是一个支持各种服务的物理基础设施。其中最成功的乃是万维网（WWW）。这是一组巨大的电子文档，被记录在组成互联网的设备之中，并通过那些为它们提供导航的超链接相互关联。这种模式有些类似于由文章、书籍、专利等组成，且经引用而相互关联的科学文献整体。

万维网的想法诞生于欧洲核子研究组织。物理学家蒂姆·伯纳斯–李（计算机科学家罗伯特·卡约后来也加入进来）于1989年提出了这一想法。伯纳斯–李设计了一个允许科学家们通过自己的计算机访问粒子物理实验所得出的大量数据的系统。让这一系统得以运转的软件并未申请专利，而是被公开发行。这一决定——就像TCP/IP的例子一样——被证明意义重大。从一开始，成千上万（之后甚至更多）的用户便对其进行了试用和改进，进而创建了各种网页和服务。短短几年之内，网络覆盖了整个世界。它的量级无人知晓，因为任何探索网络的搜索引擎（比如谷歌或雅虎）都无法归档所有网页。毕竟，这也没什么意义：几个网站一经请求就能生成新的页面。一项2005年的评估

认为，整个万维网（对于静态页面而言）的内容相当于200太字节的信息。当时，这个数据量大约相当于10个美国国会图书馆。毫无疑问，如今这一数字会比之前多出若干数量级，因为万维网的信息量增长呈指数级。

赛博空间

2001年9月11日，纽约市的基础设施经历了一次“网络灾难”，这一灾难与当天发生的人类悲剧有关且类似。在两架被劫持的飞机撞向世贸中心大厦之后不久，记录显示通话数量激增。人们试图与朋友取得联系，并与同事一道救援相关人员。手机网络很快超载，人们纷纷在曼哈顿的付费电话旁排队等候。这次袭击破坏了威瑞森[3]的总部，切断了20万条线路。美国电话电报公司的基础设施，包括那些设置在世贸中心地下室的也遭到损毁。当电话呼叫失败时，许多人便求助于互联网。但无线服务也受到影响。而袭击对经济的影响则远远超出事故地之外。纽约证券交易所耗时六天才重新恢复运转，而各种服务直到数月之后才恢复至灾难前的水平。

这场恐怖袭击表明，几乎没有任何网络能够独立存在。物理和虚拟的基础设施已嵌入公共赛博空间（cyberspace），并在其中提供能量、信息、传输和通信等服务。电网支撑着互联网，互联网存储万维网，万维网又让邮件服务、社交网络、信息网站和文件共享系统运转起来。许多活动，包括飞行管控、银行项目、应急系统和商业服务等也依赖万维网。赛博空间中某个层面的崩溃通常会以相当不可预知的方式影响到其他层面。

多个网络的相互连接在许多其他情况下也很常见。例如，影响2008年经济的流动性冲击迅速蔓延至其他许多经济网络之中。类似地，社交网络也在许多方面显示出这种特征。一个有趣的例子是真实世界的友谊与虚拟社交网络的联系相比较而显示的特征：这两个网络之间有着重要的反馈机制。细胞则是微型赛博空间：基因调控网络、蛋白质相互作用网络和代谢网络之间彼此嵌套。因此，一些科学家提出将基因组、蛋白质组、代谢组等相关概念融合为相互作用组这一综合概念。最后，生态系统可被看作是相互作用的网络组。例如，对抗和共生网络都在决定物种如何繁盛方面发挥着自己的作用。在所有这些情况下，网络为解开相互交织的复杂系统提供了有用的图景。

第四章 连接与闭合

同一个世界

2006年11月4日，德国西北部某地的一根输电线的断开引发了一场影响远及葡萄牙的雪崩式停电。人们以为这种小型断线故障的影响只限于区域层面，或至多限于国家电网范围之内。但电网的集成度已越来越高（目前，它们已在洲际层面形成了一个大型系统），因而也更为脆弱。其他基础设施的情况也一样。例如，几乎所有的机场都是相互关联的：如今，人们从任何起点出发都可以沿着经停点有限的路线抵达几乎任何目的地。互联网也是完全互联的，因为它本就是从更小的网络集成中成长起来的。

自然和社会中的网络可能在局部地区显得极为分散。细胞中的一些化学物质组仅仅在彼此之间相互作用，而不与其他任何物质发生作用。在生态系统中，某些物种群体会建立小型食物网，而与外部物种没有任何联系。而在社会系统中，某些人类群体可能也与其他人完全隔绝。然而，这种分离的群体，或者称分量，仅为极少数。在所有网络中，系统内几乎所有组成部分都会参与到一个大型的连接结构之中，这个结构被称为巨型连通分量。

例如，科学家在基于词汇联想的一个实验中，发现96%的语词止步于某个语词大群体内部。人们可以在这个词群中找到任意两个词语的关联路径，甚至像“火山”与“胃”这样差别巨大的语词之间也有关联：该实验的参与者所绘制的联想链为，“火山”——“夏威夷”——“放松”——“舒适”——“疼痛”——“胃”。一般来说，巨型连通分量在几乎所有网络中都会占到90%-95%的系统比例。我们可以列举这一事实的一些有趣后果。在性关系网络中，过去和当下的性行为能将我们与那些从未想象过或渴望有关系的人联系起来。大型的合作结构会出现在科学家的合作网络中，略去的只是少量孤僻的参与者。在公司的董事会中，连锁董事给予了绝大多数公司某种关联性。最后，在食物网中，某个给定地点的某物种所引入的污染物，会通过食物链而被带到远至地球另一端明显不相关的物种身上。

生活在一个大的相互连接的世界中并不意味着其中的任意两个节点都能够互相抵达。正如对普通道路而言，是否为单行道非常关键，

我们也必须知道边是否有向。在有向网络中，从一个节点到另外一个节点存在路径并不保证这条路反向也能行得通。狼吃羊，羊吃草，但草并不吃羊，羊也不吃狼。这种限制在巨型连通分量中创建了某种复杂的架构（图6）。例如，根据1999年的估计，若忽略边的方向，万维网有超过90%的部分由相互连接的页面组成。然而，如果我们将边的方向纳入考虑范围，万维网中彼此能够相互抵达的节点比例仅为24%，这部分构成了所谓的巨型强连通分量。其余部分则以内向分量和外向分量进行划分：前者由那些路径指向巨型强连通分量的页面组成，后者则由那些接收来自巨型强连通分量的连接的节点组成（图形由被称为管道和卷须的小型结构补充完成）。这种特征结构赋予了万维网这个巨型连通分量结构某种“蝴蝶结”的外形特征。这种复杂的结构并非局限于万维网，而是在所有有向网络中都会带着不同的分量出现。

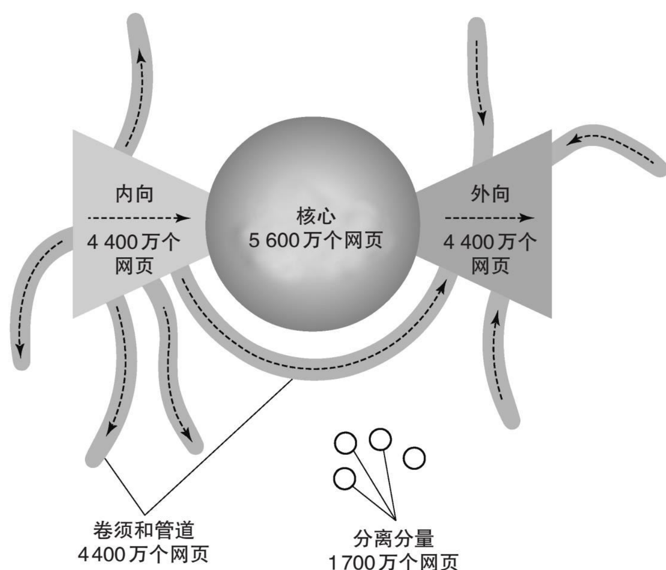


图6 像万维网这样的有向网络会显示出“蝴蝶结”结构：它由中间的巨型强连通分量、内向分量、外向分量以及各种小结构（管道、卷须以及少量分离分量）组成。图中为1999年的数据

巨型连通分量的存在——无论是否具备“蝴蝶结”结构——是一个非常显著的特征。例如，尽管万维网的巨型强连通分量的规模比系统本身的1/3还要小，但它在1999年时仍包含了5 600万个网页。如果网络非常紧密，也就是说，如果它们拥有许多冗余连接，多到能够使几乎各节点都互相连接，那么便很容易解释这一特征。但通常情况并非如此。大多数网络反而是稀疏的，即它们的连接数往往很少。以机

场网络为例：每个频繁旅行者的经验都表明，直航并不常有，抵达一些目的地需要经过中转站；活跃的机场数上千，但平均起来，每个城市仅与不到20个其他城市相互连通。大多数网络中的情况也是如此。这一情况可以由节点的平均连接数，即它们的平均度数体现出来。我们每天都创建许多网页，但每个网页的平均连接数大约为10。无数的路由器接入了互联网，但每个路由器平均只与不到3个其他路由器相连。最后，在超过5万名物理学家的样本中，人们发现合作者的平均数量仅大约为9。在大多数真实世界网络中，可能存在具有多连接的元素（事实上也确实存在），但一般说来，与此相对应的图并不紧密；相反，它们被认为是稀疏的。

这个令人困惑的矛盾——一个稀疏的网络仍旧可以很好地连接起来——已经吸引了我们在第二章提到的匈牙利数学家保罗·埃尔德什和阿尔弗雷德·雷尼的注意。他们通过生成这些图的不同随机图解开了这个困惑。在每个随机图中，他们都改变了边的密度。他们以非常低的密度为起点：平均每个节点少于一条边。人们自然会期待，随着边的密度的增加，越来越多的节点将会彼此连接。但埃尔德什和雷尼所发现的却是一个相当突然的转变：一些分离分量突然合并为一个大的分量，后者几乎包含了所有节点。突变发生在边达到一个特定的临界密度值时：当每个节点连线的平均数（即平均度数）大于一时，巨型连通分量便会突然出现。这个结果意味着网络展现出某种十分特别的节约特征（这种特征是其无序结构所固有的）：甚至在节点之间随机分布的少量的边，都足以产生能够吸收几乎所有元素的大型结构。

近在咫尺

1994年初，奥尔布赖特学院的三名学生——克雷格·法斯、布莱恩·特特尔、迈克·吉内利——在一场暴风雪天气里看电视。据他们说，他们注意到电视上预告了凯文·贝肯的下一部电影，但他也同时出现在了许多不同的电影中。于是他们开始历数在那些电影中与他合作的大量演员。贝肯有点类似于“娱乐世界的中心”的想法便开始传播，进而变得尽人皆知了，基于此甚至出现了一个网页，即“凯文·贝肯指南”：这个搜索引擎提供了贝肯和人们可能会搜索的任何其他演员之间的关系。值得注意的是，如果人们输入来自早先商业电影中某位西班牙演员的名字，比如帕科·马丁内斯·索里亚，该指南网页就会发现一个十分紧密的联系：马丁内斯与路易斯·因杜尼一起演过《在西班牙度假》（*Veraneo en España*）这部电影；后者和伊莱·瓦拉赫一起出演过电影《先开枪，再提问》（*Il Bianco, il Giallo e il Nero*）；而瓦拉赫又与凯文·贝肯一起出演过《神秘河》（*Mystic River*）。这种与贝肯相关的合作链几乎能在人们所能想到的任何演员身上发现。

这个令人惊讶的特征让人想起一场科学家参与的游戏。随机图专家保罗·埃尔德什乃是20世纪杰出的数学家。科学家们衡量了自己与埃尔德什的合作以作为自己荣誉的象征：那些与他共同撰写过一篇论文的人的埃尔德什数记为1；那些埃尔德什合著者的合著者的埃尔德什数记为2；而埃尔德什合著者的合著者的合著者的埃尔德什数记为3；以此类推。然而，只有最小的埃尔德什数字才是骄傲的真正来源：超过500位科学家乃埃尔德什的直接合著者；而与这些核心的合著者合作过的科学家有几千人。最后，数以万计的人拥有埃尔德什数3（G.C.[4]就是其中之；M.C.则拥有埃尔德什数4），所以这并非什么特别的荣誉。在一任何领域几乎都没有科学家拥有超过13的埃尔德什数。

完全出乎人们的意料，这些引人注目的结果并非贝肯或埃尔德什特有。前者并非娱乐界的中心人物；后者也并非数学的枢纽。如果在任何其他演员或科学家身上重复同样的计算，也会发现类似的结果：非常短的链条关联着相隔甚远的个体。这一事实为一次十分普通的鸡尾酒会经历提供了某种有趣的洞见：

当你与一位陌生人聊天的时候，突然发现他或她是你妻子的同学，或者你兄弟的网球搭档，抑或你朋友的邻居，等等。这种发现通常让人们惊呼（“世界真小！”），但可能它并非那般不寻常。

社会系统似乎非常紧密地联系在一起：在一个足够大的陌生人群中，找到一对由十分短的关系链关联起来的人并非不可能。

六度分隔理论

1967年，美国心理学家斯坦利·米尔格拉姆进行了一系列值得铭记的实验。在杰弗里·特拉弗斯的协作下，米尔格拉姆向中西部（堪萨斯州和内布拉斯加州）的一些居民随机寄送了数十封信件。在信中，他请求这些人将信件转发至马萨诸塞州的某人（剑桥[5]的神学学生的妻子，或者波士顿的一位股票经纪人），而米尔格拉姆并未提供相关地址。如果这些居民并不认识收件人，米尔格拉姆则建议他们将信件转发至他们所认识的可能由于某种原因而与收件人“相熟”的人。人们在每次转发信件之时，都必须抄送一份给米尔格拉姆，以便他能够追踪信息的传递路径。在拥有数亿人口的国家内，依靠本质上为口口相传的方式找到某人似乎是不可能的。然而，几天后，收件人开始收到原始信件。这些信件仅仅经过了一位中间人。几周之后实验宣告完成，大约1/3的信件已经抵达其目的地：没有一封信件被转发超过10次，而平均邮寄次数为6次。

该实验激发了科学界的热情。1950年代，数学家曼弗雷德·科亨和政治学家伊锡尔·德·索拉·普尔便猜测人类之间的联系可能比预想的更加紧密。他们问道：如果从人群中随机抽取两人，他们认识彼此的概率有多大？通俗地讲，需要多长的相识链条才能将他们相互关联起来？在一篇最终于1978年发表的广为流传的论文中，他们提出了一个数学模型，表明在诸如美国这样的人口规模中，让人意想不到的是有很大一部分人两两之间都可以通过少量的中间人链条而彼此连接。米尔格拉姆的实验则是对此二人直觉的检验。这个发现有很强的甚至超出学术界的影响。六度分隔理论这个术语中的数字来源于米尔格拉姆的实验发现，并成为其结果的一个通俗的表达。1990年，剧作家约翰·瓜雷将其作为某喜剧的标题，该剧中一个魅力人物躲避人群，并声称自己是演员西德尼·波蒂埃的儿子。以下便是瓜雷表达米尔格拉姆实验结果的方式：

> 我曾在某处读到过，这个星球上的每个人和其他人之间都仅隔着六个人。这就是六度分隔理论。在我们和这颗星球上的其他每个人之间。美国总统。威尼斯的贡多拉船夫。还可以继续写下名字。［……］不仅仅是名人。而是任何人。雨林中的当地人。火地岛人。爱斯基摩人。我通过六个人与这颗星球上的每个人紧密地联系在一起。

小世界

“爱虫”（*Iloveyou*）病毒是史上最具传染性的计算机病毒之一。自2000年5月出现之后，它便感染了世界各地数千万的计算机，造成主要用于清除它的损失达数十亿欧元。“爱虫”病毒以电子邮件的形式传播，它伪装在一个酷似情书的附件中。当附件被打开，病毒便会感染计算机，并将自己转发至该计算机的地址簿中所含的电邮地址中。经过仅仅几次这样的复制，该病毒便感染了大量设备。正如前面某段落所描绘的社交网络一样，人们或许这样谈论病毒传播的计算机网络：“世界真小！”仅仅数次连接便可访问大量计算机。彼此明显相隔甚远的计算机结果仅通过较短的关系链便相互连接。

在现实中，作为米尔格拉姆实验主要结果的这种小世界属性，存在于所有网络中。互联网由数十万个路由器组成，但信息包仅通过大约10次“跳跃”便足以从其中一个路由器抵达其他任何一个。两者之间可能相隔数千公里，但距离并不重要：重要的是必经的连接数，而这个数字往往很小。这就是为何信息能以非凡的速度在地球上传播的原因。另外一个例子是万维网：它由数十亿页面构成，但科学家发现，通常20次鼠标点击便足以连接其中任何两个页面。一般而言，在一只

秀丽隐杆线虫的大脑中，任何一对神经元之间的距离都少于“三度分隔”。而在连接世界上所有国家的进出口网络中，无法找到被超过两条联系链条分隔的国家。这样的例子在其他很多事例中数不胜数。

小世界属性包含任意两节点之间的平均距离（被测定为关联它们的最短路径）十分小这一事实。给定某网络中的一个节点（比如在合著者网络中的埃尔德什），少数节点与它十分接近（即直接的合著者），与其相隔甚远的仍是少数（即那些持有很大埃尔德什数的科学家）：多数人则位于平均——以及很短——距离之内。这一点适用于所有网络：从一个特定节点开始，几乎所有的节点都仅与其相隔寥寥几步；特定范围内的节点数则随着距离的增加以指数方式快速增加。相同现象的另一种解释方式（科学家们通常的说明方式）如下：即便我们向某个网络中增加了许多节点，平均距离也不会增加太多；人们不得不将网络大小增加几个数量级后，才能注意到通往新节点的路径变长了（一点点）。

小世界属性对许多网络现象至关重要。新皮质中较短的突触距离对其功能可能非常关键：一些研究表明，神经退行性疾病如阿尔茨海默症便说明大脑中的小世界属性大量丢失。而性关系网络中的短程关系则表明，解释性传播疾病中的风险群体这一概念时必须小心翼翼：事实上每个人距离那些被感染者都很近。网络这种有效传播病毒的能力也有其建设性用途。采用病毒式营销策略的首批案例之一是微软1996年推出的电子邮箱服务（Hotmail）的全球性扩散。购买免费微软电邮地址的人须同意在他们的信件中附加一个链接，该链接允许收件人相应地打开一个免费地址。微软电邮在通信公司中增长最为迅速，它俘获了数千万用户，部分原因在于这个策略巧妙地利用了电子邮件网络的小世界属性。

捷径

小世界属性是网络所固有的。即便完全随机的埃尔德什–雷尼图也表现出这种性质。相比之下，普通的网格并不会展现这一特征。如果互联网是一个棋盘状晶格，则任意两个路由器之间的平均距离将达1 000步的级别，而网络本身也会慢很多：不会有快速的网页浏览，也没有即时的电子邮件。如果科学合作网络是一个网格，则埃尔德什仅会有中等数量的合著者；尽管后者的合作者数量会大些，但规模仍旧适中，以此类推：特定距离内的个体数量不会呈指数级增长，而是慢许多。如果神经网络是一个晶格，则增加神经元的数量（比如，由于大脑的自然发育）将显著增加新皮层内的平均传递距离：矛盾的是，大脑的发育会让人变得不那么聪明（一些年轻读者可能会同意这

一点）。

那又是什么使得网络与网格有所不同？网络中表现出的小世界属性为何又在晶格中缺乏？1998年，物理学家邓肯·沃茨和数学家史蒂文·斯托加茨试图回答这些问题。他们以思考十分简单且规则的结构为出发点。这个结构为节点构成的圆形，其中每个节点都与其最近和次近的邻点相连（图7左）。这种结构可能代表了一组偏远的村庄，其中每个村庄都与其邻近村庄交换货物，偶尔也会与邻村的邻村交换。在这种规则的结构中，产品从生产者那里传递至遥远村庄的消费者处可能要很长一段距离。

接着，沃茨和斯托加茨引进一个类似规则使得两个遥远村庄之间开放路径。在实际操作中，他们切掉了原始结构的某条连线，并将其与其他节点重新随机连接。顷刻间，一个村庄的居民便能与之前遥远的地区而非其邻村交换货物了。然而，从整体上看，仅有几个村庄受到了这种变化的影响，圆圈中的一些区域彼此之间仍旧相隔甚远。人们可通过计算重连之后节点之间的平均距离看出这一点：引进新的捷径之后平均距离仅稍微变短。这时，两位科学家开放了更多的“路径”（图7右）。每次重新连接之后，他们都会计算节点之间的平均距离：他们发现，在重新增加仅仅几处连线之后，平均距离明显缩短。少量捷径便足以大大缩短系统中所有元素彼此之间的距离。将连接结构转变为小世界的关键因素便是少量无序的出现。没有哪个真实的网络能呈现出有序的元素排列。相反，其中总存在一些“不合群”的连接。正是由于这些无序连接的存在，网络方为小世界。

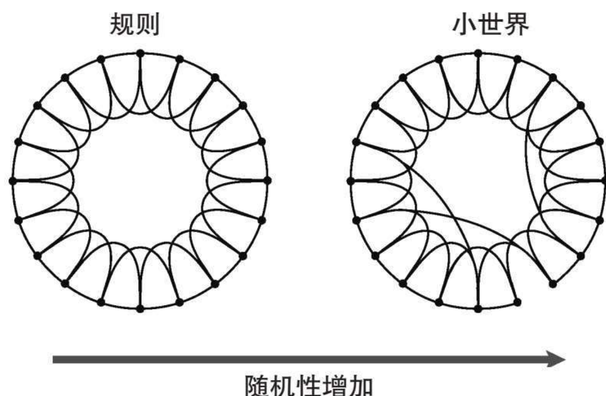


图7 在小世界网络模型中，引入无序后规则晶格便转换为网络，相应地，节点之间的距离也缩短了：小世界属性由此产生

这些捷径在某些网络中很容易识别。例如，一1858年第个跨大西

洋的电报电缆将欧洲和美国连接了起来。这个长达数千公里、重达数百吨的奇迹是儒勒·凡尔纳的《海底两万里》中潜水艇造访过的海洋奇迹之一。如今，数条跨洋电缆让信息能够在世界范围内即时传递。在语词网络中，一词多义乃捷径的主要来源之一。比如，“pupil”一词连接着两个语义区，一个是“教学”（“pupil”作为“学生”），另一个是“视觉”（“pupil”作为“瞳孔”）。在社交网络中，格兰诺维特的弱连带概念——将不相关群体连接起来的联系——至少可以与沃茨和斯托加茨的捷径概念部分对应。

捷径在许多其他情况下也是引发小世界属性的原因之一。然而，我们不时也能找到另外可能的解释。例如，世界贸易网络中的超短距离则源于它是少数高度紧密（非稀疏）的网络之一这个事实：一个国家的贸易伙伴平均数量一般可与这个网络里的国家总数相当，这意味着每个国家都与其他大多数国家相互进行贸易。在食物网中，其他机制也能促成小世界属性。基位物种从阳光和周围环境处获得能量和物质，但它们的捕食者通常仅能摄取它们体内资源的10%，而在次第展开的每一步捕食中都是如此：如果食物链太长，顶端的捕食者便无法摄取足以维持其生存的资源。

无论最初怎样，当系统有了某种图形结构之时，小世界属性便成为需要考虑的重要特征。网络方法提供了这种系统的某种醒目景象：首先，系统的元素是某个大世界的一部分，其中几乎每个节点都具有彼此连接的路径；其次，这些路径非常短。这种交织的结构对理解从艾滋病到断电再到信息传播等广泛存在的现象而言至关重要。

第五章 超级连接器

枢纽

在其令人印象深刻的“六度分隔理论”诸实验中，斯坦利·米尔格拉姆做了一项很久之后人们才能充分理解的观察。其中一个实验里，美国心理学家请一些随机挑选的内布拉斯加州公民转发一封信至马萨诸塞州的某股票经纪人。如果他们不认识收件人，就把信件寄送至那些他们认为与该经纪人更为接近的人。除了大部分信件平均仅用6步便抵达收件人处这一事实之外，米尔格拉姆还观察到，1/4的信件经由同一个源头传递至收件人，即该股票经纪人的服装商朋友，米尔格拉姆称之为雅各布先生。这一结果令人相当费解：为何如此之多指向该股票经纪人的关系路径都要经过这位先生？

经常坐飞机的人对类似现象都很熟悉。希思罗、法兰克福或纽约肯尼迪等机场对于环球旅行者而言都耳熟能详：无论目的地是何处，飞机都很可能会停靠这些机场。航空杂志常常附有世界地图，纵横交错的长线标出飞行路线：其中许多航线都会途经伦敦、法兰克福以及纽约等地，或将其作为终点。像这样的机场被称为枢纽，它们承载了整个航空交通的大部分运力。我们很容易得出，雅各布先生在社交网络中的位置与这些大型机场在空中交通中的地位相同。很可能，雅各布就是一个社交关系枢纽：他的许多联系人将其与其他一些人相互连接，所以很自然地，许多信件都要经他发出。

米尔格拉姆另一个引人注目的观察是，剩余大部分信件经由另外两人寄达：琼斯先生和布朗先生。借用空中交通这个比喻，这两人最可能是社交网络中“一般规模的机场”（如马德里或米兰等机场）。而剩余那些并非来自雅各布、琼斯或布朗的信件则途经社交网络中的较小“机场”（如吉罗纳或奥尔比亚等机场）。

这些枢纽的存在并不局限于该股票经纪人的社交网络或者机场网络。在许多其他系统以图表示后，其中也能看到类似的高度连接顶点或超级连接器。在许多网络中，人们都能看到“赢者通吃”的趋势：少数几个节点能吸引大多数连接，而余下的大部分节点将不得不共享剩余连接。现代分析表明，莫扎特创作的唐·乔万尼（此人引诱了2 065名女性，根据达蓬特的剧本：“……在意大利640名，德国231名，法

国100名，土耳其91名，而西班牙境内已达1 003名……”）这一人物并非夸张：在性行为网络中与他人连接最多的个人的性行为可达数千次。在一些数据集中，此类人中的一部分涉及性交易。自然地，这些与他人高度关联的个体最需要预防性传播疾病的感染。“9·11”恐怖袭击在纽约发生后不久，人们发现了超级连接器的又一例证：管理顾问瓦尔迪斯·克雷布斯画出了恐怖分子的社交网络简图，他发现这场阴谋的领导者之一穆罕默德·阿塔是连接最多的节点，即此人为该社交网络的枢纽。在科学合作网络中，我们也能发现那些与大量同行合作的关键人物：保罗·埃尔德什便是其中之一。

除了社交网络，超级连接器还存在于各种各样的网络中。互联网中的某些路由器具有数千条连接：这比一般的路由器多出几千条，后者仅有少数几条连接。汇接机房乃大型设施——通常为装满电缆的建筑物——数百个互联网服务商可经由它彼此连接：这些设施只要有一处出现故障就能导致整个区域（大到一个国家）失去网络连接。大型报纸的网站吸引了大量来自其他网站、博客和社交网络的链接。而在食物网中，位于顶端的物种捕食其他大量物种。最后，语词网络中的枢纽则为模棱两可或多义的语词：比如“arms”，它在英语中既指身体四肢也指武器，因而与更大的语义场或同义词域相连。

枢纽也存在于细胞内的网络中。在基因调控网络里，单个基因可以控制大部分剩余基因组的表达：在某种细菌（新月柄杆菌）中，一个调节因子（CtrA基因）便可控制26%的细胞周期调节基因。p53分子是蛋白质相互作用网络的超级连接器：与该蛋白质相关的基因是强大的肿瘤抑制基因，它会在大量的肿瘤里发生突变。代谢网络的枢纽很明显是ATP分子（腺苷三磷酸）：它在大量的生物化学反应中起到能量载体的作用。

巨人、侏儒和网络

> 从非利士营中出来一个讨战的名叫歌利亚的迦特人，身高六肘零一虎口。［……］头戴铜盔，身穿铠甲，甲重五千舍客勒［……］铜戟枪杆粗如织布的机轴，铁枪头重六百舍客勒。（《撒母耳记上》，17：4-7）

根据《圣经》的《撒母耳记》，以色列人必须等待40天才有人敢于面对歌利亚那般强壮的人：随后大卫这个勇敢无畏的男孩进来了，他最终击败敌人。歌利亚并非寻常敌人：“六肘零一虎口”的身高相当于大约3米高，而根据历史学家的推测，他身上“五千舍客勒”的铠甲也重达60-90公斤。

古代计量换算为现代计量时并不是很精确；而且，《圣经》的记述很可能是象征性的。然而，歌利亚的身高不是完全没有可能。根据《吉尼斯世界纪录》，有记录的最高者为一个名叫罗伯特·瓦德洛的美国人，此人身高2.75米。与歌利亚有着特制的盔甲和适合其身高的矛不同，特别高的人周围的物体与之相比都太过短小：椅子不舒服，天花板太低，他们还需要穿特制的鞋子和衣服。

他们的问题根源在于，身体尺寸是一种同质量值。进入电影院的人有着不同的身高，但所有的座椅都一样：有些人觉得椅子大了，其他人觉得小了，但一般而言，他们都觉得还算舒服。身体尺寸不会偏离平均尺寸太多。很高（或很矮）的人非常罕见，越高（或越矮）的人则越少见。几乎每个人都认识1.9米高的人，但仅有少数人认识2米高的人，而几乎没有人认识身高2.5米的人。人们在其他一些特征上也具有同质量值。例如，人们的智商测试结果多数时候接近平均水平，而偏差——无论往上还是向下——则较为罕见。人们的行为方式也十分同质化。比如，司机可能多少都有些莽撞，但经测量，多数时候他们在高速公路上的行驶速度都非常接近平均水平。

然而，同质性并非金科玉律。例如，一个人的朋友数量是极度多变的。根据《吉尼斯世界纪录》，瓦德洛的身高仅是最矮的人的五倍，后者名为钱德拉·巴哈杜尔·唐吉，身高55厘米。相比之下，最友好的人（即社交网络中的枢纽）所交的朋友数比那些仅与很少人交往的极度害羞之人要多出数十上百个。如果将虚拟社交网络中的联系人也算入一个人的朋友之中，那么这些网络的枢纽人物会比那些不善交往的人多出数百个朋友。人们的身高属于同质量值，但社交关系的数量却是异质的。

如果人的身高反映了他们社交关系的数量，那么，像瓦德洛这么高的人不会进入任何世界纪录。社交关系中会有比矮子高出几百倍的人：身高超过两公里的“社交巨人”会行走在社交之路上。更有趣的是，这些巨人在普遍矮小的人群中并不会是惊人的例外。侏儒和巨人之间的所有中间高度将由另一些人代表：自然，高度越高，人数越少；然而，这个想象世界里的高个子数量不会像在现实世界中那样迅速减少。换句话说，越高越少，但也不至于像在现实世界中那般稀有。

在这个想象世界中，座椅制造商的业务难度会增大许多，因为没法制造一个适合每个人身体尺寸的座位。而在现实世界里，如果想制造座椅、分析智商测试或预测自驾旅程的时长，我们会考虑平均身高、智商或行车速度。但为了理解社会关系，平均的概念就显得无用了。身体尺寸、智商、行车速度以及其他量值都具备特征尺度，即大

多数情况下的平均值都是对我们所发现的实际值的大致预测。相比之下，社交关系并不具备这种尺度。如果去敲一个陌生邻居的门，你预计看见之人的身高会在一个合理的范围之内，而你的猜测多数时候是准确的。但我们几乎不可能提前猜测此人朋友数量的多寡以及具体数字。某个城镇的平均人际关系数量仅能让我们了解该地区社交网络的疏密程度。但我们无法据此对每个个人做出任何合理的预测。具备这种特征的系统被认为是无标度或标度不变的，意为该系统并不具备特征尺度。这句话还可以这样表述，相较于平均值，个体波动太大，以至于我们无法做出正确的预测。

肥尾效应

一般而言，具备异质连接性的网络都会有一组清晰的中心。当图很小时，我们很容易发现其内部连接是同质还是异质的（图8）。在第一种情况下，所有节点多少具备相同的连接性，而在后者中则很容易发现少量枢纽节点。但是，当被研究的网络非常大（如互联网、万维网、代谢网络以及许多其他网络）时，事情就没那么简单了。幸运的是，数学提供了一种方法来确定一种量值是同质还是异质。

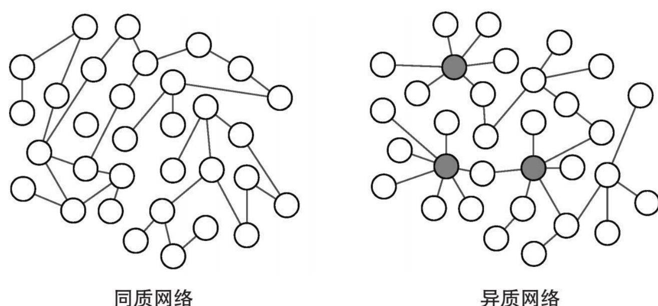


图8 与存在高度连接节点（枢纽节点）的异质网络（右）相比，同质网络（左）中所有节点的度数大致相同

我们以同质量值为起点，比如人的身高。为了研究某班学生的身高，我们可以按照以下方法操作。首先，让那些身高在1.50到1.55米之间的学生排成一列：他们可能人数不多。然后，让那些身高在1.55到1.60米之间的学生平行地排成一列：这些人的数量会多些，队伍也会长些。接下来的一列为1.60到1.65米的学生：更多的人会出现在这一列。然后，每一列身高增加5厘米（图9左）。最后，这些列的轮廓将构成钟形曲线的形状：

学生的数量随着身高的增加而增加，然后在平均值附近达到峰

值，接着开始下降。很高和很矮的学生都较少，大部分处于中间范围。这条曲线提供了学生的身高分布。

现在，我们来考虑这些学生的社交关系数量。这时，每一列分别对应0到20个朋友，20到40个朋友，40到60个朋友，以此类推。该过程的结果提供了社交网络节点的连接性分布，即图的度数分布。这个结果图与身高图的情况十分不同（图9右）。首先，图中的列会更多，因为有的人的朋友数量会成百上千。多数人的联系人为几十个，但由此产生的分布将具有“肥尾效应”。换言之，分布图的长尾或者说是“厚尾”将明显向右偏斜。从数学角度讲，度数分布的形状可通过幂律得到很好的描述。

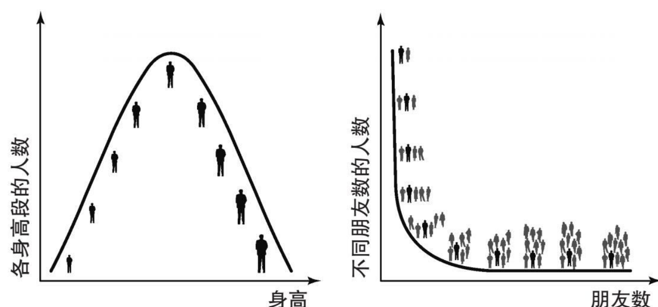


图9 人们的身高为同质量值，呈钟形曲线分布（左），而人们的朋友数量则为异质量值，呈幂律分布（右）

在同质网络中，度数分布是类似于前述学生身高的钟形曲线，而在异质网络里，度数分布则遵循幂律，类似于朋友数量的分布图。幂律意味着异质网络中存在着比同质网络更多的枢纽节点（以及更多的连接数）。此外，枢纽节点并不是单独的例外：与连接数较少的网络相比，连接数较多的网络中有着完整的节点层级结构，每个节点都构成了一个枢纽。再拿身高和朋友数量来说。世界上身高1.50米的人可能有数百万；然而，如果我们将这一高度翻倍（即3米），如此高度的人则少得多，很可能没有。另一方面，数千万人在其社交网络中有比如说20位朋友。如果我们将这个数字加倍（40位朋友），拥有这个朋友数的人数则会少些（比如比加倍之前人数减少了1/4），但仍有数百万。我们可将这个数字多次加倍，而对应的人数则每次减少约1/4（实际减少的速度取决于幂律的斜率）。这便解释了比如琼斯先生和布朗先生在米尔格拉姆实验中的作用：虽然雅各布是股票经纪人社交网络中的最大枢纽，而琼斯和布朗是更小的枢纽，但与他们联系的人依然很多。

查看度数分布是检查网络是否为异质结构的最佳方法：如果度数

分布呈肥尾，则该网络将有多个枢纽且为异质结构。人们从未发现某种数学上完美的幂律，因为这将意味着存在拥有无限连接数的枢纽。然而，不存在无限大的真实网络：这就是为何度数分布的肥尾总有一个度数最大值上限的原因。实际上，枢纽的大小会受到连接累积的各种成本的限制：例如，由于神经元的物理结构，神经元无法累积任意数量的连接。在专业协作网络中，时间起着某种作用：连接数无法无限累积，因为个体的事业（或生命）会在某个时刻终结。所有这些和其他因素都反映在度数分布的形状上。尽管如此，严重偏斜的肥尾状度数分布仍是异质网络的清楚信号，即便它从来都不是一个完美的幂律。

在解释枢纽和肥尾的含义时必须小心谨慎。例如，一些人类学家认为，一种叫邓巴数字的量值限制了人们的社交关系数量。根据这一假设，稳定的社交关系数量不能超过150这个数字太多。人类学家罗宾·邓巴在发现灵长类动物和人类的大脑皮质某部分的大小可能与它们的社会群组规模相关的证据之后，于1992年提出这一假设。如果这一假设为真，那又如何解释人们在许多社交网络中发现的有着上千联系人的社交枢纽呢？

一些科学家认为，这便是“披萨送货员问题”的实例。披萨送货员在自己的手机上会接到许多电话，但只有极少的一部分来自其真正的朋友；其余则为客户。根据这种想法，呈现在社交网络度数分布图肥尾处的多数连接都是泛泛之交。然而，这还得看人们究竟想研究什么问题。例如，如果披萨送货员得了流感，流行病学家只会关心有多少人（不管是不是朋友）曾与他有过接触。

另一方面，并非所有网络都是异质结构。尽管小世界属性是网络结构所固有的，但不是所有的网络中都会出现枢纽。例如，电网通常就仅有少量枢纽。还有一些食物网、线虫的神经网络以及世界贸易网络等都很少有枢纽存在。

最后，人们在一些有向网络中发现了一个有趣的情况，正如在多数基因调控网络中发现的那样。如果基因A调控基因B，则箭头从A指向B，但B不是一定要指向A。出度分布（即朝外箭头数的分布）通常是肥尾状：少数基因调控大部分基因组。然而，入度分布（即朝内箭头数的分布）则均匀得多：少数其他基因调控某个基因。异质性在许多网络中都广泛存在，但当我们处理未知系统时，在检验之前我们不要理所当然地认为它就是异质网络。

自组织的标志

异质性和特征尺度的缺乏可能是无序的极好标志。推论如下。许多网络（比如互联网或社交网络）是在没有任何蓝图或监督的情况下成长起来的。因此，网络中的每个节点都遵循其自身的标准，并表现出彼此完全不同和不协调的行为。这些节点十分混乱，以至于它们很容易就被某种总体无序的过程所同化。因此，随机图应该成为这些网络的良好模型。这种推理似乎有效，但若深入检验，一些问题便出现了。最明显的是，随机网络根本不是异质结构。相反，它们的度数分布为钟形结构，这表明所有节点都有着大致相同的度数。随机关联节点的过程便是如此，即每个节点最终都有着同样的度数。更确切地说，度数具有特征尺度，且在平均值附近小幅波动。与许多真实的网络相比，随机网络中不存在枢纽节点。

其结果是，随机网络为小世界，而异质网络则是超小世界。也就是说，异质网络顶点间的距离小于其随机对照网络中的相应距离。如果人们在一个随机网络中加入一定数量的枢纽节点（因而使其更具异质结构），则节点之间的距离会变小。相反，如果人们将一个异质网络随机化（即用相同数量的点和边建构一个网络，但让边随机分布），枢纽节点便会消失，节点之间的平均距离也会变大。这表明枢纽节点承担了这些网络中的大多数连接：这些连接中的大部分正是从这些少数的超级连接点中产生的。

更重要的是，随机网络为同质结构这一事实意味着异质并不完全等同于无序。如前述随机网络那般的无序过程并不会产生那些在现实网络中所发现的异质连接。相反，异质性可能恰恰来自其对立面，即来自某种规则的有序行为。

这一点十分令人困惑，因为许多网络都不是设计的结果，也并非是在自上而下的严格监管中发展起来的。像电网或道路网络等少数网络由政治和技术权威控制，但大多数网络则无人监管。例如，互联网由地方一级的网络管理员控制，并且也受到技术、经济和地理特征的限制。然而，其大规模的结构在很大程度上则是未经筹划的：互联网非常类似于一项全球规模的实验，其中没人提出整体结构，而是靠无数代理人的行动将其建立。生物网络是一个更为清晰的例子：它没有设计者，仅有进化提供修补效果。社交网络、政治、金钱、宗教、语言和文化都会对个人之间的关系产生影响，但是当个人关系存在自由空间的时候，这些网络的形成便不再是严格规划的。在所有这些情况下，系统的整体组织产生于其组成部分的集体行为，即某种自下而上的自组织过程。这个过程可以解释为何许多网络即便没有规划，却仍然显示出异质性这样明显的有序标志。

异质性并非网络系统所独有。例如，地震强度便有肥尾状的分布

特征，而如果人们绘制出地震频率与其强度的对比图，它便会呈现出很好的幂律分布。“平均强度的地震”并不存在，其中有很大的多变性，从无法察觉的震动到大规模的灾难。另一个例子是城市的规模：其范围从中国最大的大都会到托斯卡纳的小镇不等。还有就是收入分配：20世纪初，经济学家维尔弗雷多·帕累托指出意大利80%的土地掌握在20%的人手中。所有经济体中都存在不同程度的此种不均衡。

所有这些例子与网络共有一个基本特征：它们都是复杂且大体上无监管进程的结果。异质性并不等同于随机性。相反，异质性可能成为某种隐藏有序结构的标志，它并非某种自上而下的计划，而是由系统的所有元素共同作用而产生。这个特征在大范围不同网络中的出现表明，在许多这种网络中，可能有着某种共同的底层机制在起作用。了解这种自组织秩序的起源则是网络科学面临的核心挑战之一。

第六章 网络的涌现

永恒的变化

到20世纪90年代，互联网很大程度上还是一个未知的领域。尽管它已经成为通信、贸易和运输的关键基础设施，但没人清楚地知道其整体架构。管理员控制着本地网络，但他们并不清楚互联网的大规模结构。此外，互联网呈爆炸式增长：从70年代初的几十台机器到上千万台，且增长前景更为广阔。到90年代末，像康柏电脑公司（Compaq）和互联网数据分析合作协会这样的组织启动了一系列绘图计划，旨在探索互联网并描绘其全球布局。康柏公司的计划名为“墨卡托”，以此纪念在16世纪绘制了最为重要的世界地图之一的地理学家，该地图将刚被发现的美洲大陆包含在内——而互联网被认为是一个尚待探索的“新世界”。多亏了这些以及其他项目，互联网地图才得以成为可能，它的增长现已受到监控。

但互联网的动态并未停止：此时此刻，路由器、计算机、电缆和卫星连接都仍在不断被添加或移除，这带来了持续增长的净效应。尽管缺乏规划，但互联网的增长过程也并非完全随机。相反，互联网是一种高度有序和高效的结构。这种涌现的秩序一定是构建网络之个体行为的某种规则性的后果。小规模的机制也必定存在，它们经由大量的交互迭代过程，最终产生了一种在大规模层面井然有序的结构。而解开隐含于网络结构形成过程之下的基本原理，对理解这些网络结构的自组织进程而言至关重要。

甚至那些看起来完全静止的网络，实际上也在进行着某种动态过程。细胞中的基因、蛋白质和代谢物所构成的网络，大脑中的神经网络以及生态系统中的物种网络等等，它们似乎都固定不变：基因调控、代谢途径、神经元连接或猎物–捕食者关系都相对稳定。然而，细胞内的网络在每个个体的发育期间呈爆炸式增长，并且随着生物体的老化及其对环境的应变过程而不断变化。大脑的可塑性在整个生命过程中可能会下降，但它并不会完全消失。物种灭绝或新物种入侵则会彻底重塑食物网。此外，所有生物网络在长期内都会因为自然选择的作用而发生改变。类似的事情还发生在其他明显的静态网络中，例如电网或语言网络。电网建成之后，由于事故和技术发展，它们会慢

慢调整自身的形状。语词网络则随着说话者的改变而改变，并且，随着语言整体的演变，新词和新的语义关系也会被引入。

在其他网络中，变化主要集中在连接上，而顶点集合几乎不可改变。例如，某一国家的银行每天都会在银行同业网络中建立不同的借款模式。偶尔，该网络的顶点集合也能发生改变，比如在银行破产或者新的银行出现在市场上之时。然而，这种变化发生在比交易（即边的重新排列）更长的时间段内：在给定的一天中，网络中的变化主要与边相关。类似的事情不仅发生在股票间的价格相关性网络中（相关性的变化比实际的股票组频繁得多），也发生在世界贸易网络中（世界各国经济关系的变化比新的国家通过分离或联合而建立更为普遍），还有机场网络中（不管在哪一年，都仅有少数新机场开张，而大量航线则处于变化之中）。

完全相反的情况则出现在另外一组网络中：在这些图中，新节点被不断添加进来，而这一进程比连线的重新排列重要得多。这种情况最确切的实例是科学论文中的引用网络。引用较早论文的新论文每天都会出现，一旦发表，相应的引用关系便无法再次更改。

最后，在一些网络中，添加（或减少）节点和连线的动力会以相同的速度发生，并让位于一个相当复杂的过程。万维网以创建和删除新网页和新超级链接的方式不断更新。在维基百科这样的特定网站中，每天都会有新文章和文章间的新链接被创建出来。

可能的网络动力的范围十分广泛：网络科学家已经制定了各种衡量标准、数学模型和计算机模拟，以掌握这一过程背后的基本机制，以期理解这些网络的自组织原理。

富者更富

20世纪60年代初，社会学家哈丽雅特·朱克曼采访了一批获得诺贝尔奖的科学家，其目的是找到这些人工作方式中的独特之处，以及他们研究工作成功的秘诀。她在这些诺奖获得者的回答中发现了一个反复出现的主题。一位诺贝尔物理学奖获得者说：“在授予荣誉这件事上这个世界很奇怪。它往往将荣誉给予那些（已经）出名的人。”一位诺贝尔化学奖获得者补充道：“当人们在报纸上看见我的名字，他们往往会记住它，而忽略其他人名。”而一位生理学和医学奖获得者则具体谈道：“［当你在读一篇科学文献时］你常常注意到你熟悉的名字。即便排到最后，它也会让人印象深刻［……］你会记住它，而非长长的著者名单。”“最出名的人会获得更多的声誉，在程度

上毫无节制的声誉。”物理学奖获得者如此总结道。

这些观察促使另一位社会学家（朱克曼后来的丈夫）罗伯特·默顿于1968年提出了一个出色的定理。默顿提出，科学处于马太效应这种社会机制的影响之下。这个名字来自《马太福音》：

> 凡有的，还要加给他，叫他有余；没有的，连他所有的也要夺过来。（《马太福音》，25：29）

依照默顿的观点，这种机制在奖项、资助、关注度、声望等的分配中起着一定的作用。拥有大量此类资产的科学家可以轻松获得更多的类似资产。另一方面，那些缺乏者想要获取或保持这些资产就十分困难。

1976年，物理学家和科学史家德里克·德·索拉·普莱斯发现了支持这种观点的量化证据。普莱斯分析了大量因相互引用而彼此关联的科学文献。他发现，在特定时期内被大量引用的文献往往也会在之后获得更多引用，而那些一开始仅被少量引用的文献后来也未增加多少被引数。普莱斯从数学上证明了这个简单的法则能够解释高引文献（即引用网络的枢纽节点）出现的原因。更确切地说，他证明了这种机制能够解释为何每篇文章的引用数量分布显示出了幂律的肥尾特征。

普莱斯的模型是统计学家乔治·尤德里·尤尔和社会科学家赫伯特·A.西蒙早前开发的数学模型的变体，但是人们直到1999年才明白，这种机制可以解释网络枢纽、异质性、标度不变性在不同网络中整体以肥尾分布出现的原因。该模型的实现归功于物理学家奥尔贝特–拉斯洛·巴拉巴西和雷卡·奥尔贝特。巴拉巴西和奥尔贝特提出了网络增长的数学模型。他们设想了一个始于少量顶点（甚至仅有两三个）随机连接的图形。新的节点以稳定的速度加入到这个原初的核心中，每个节点都带有给定数量的连接。一个简单的规则确定了新节点是如何连接的：新节点会偏好已有许多连接的旧节点。这种机制又称优先连接（图10）。原则上，新节点可以加入任何旧节点之上，但旧节点的度数越高，它吸引新节点的概率便越高。偶尔，连接较少的节点也会接入新的连接，但大多数时候却是枢纽节点更具吸引力。

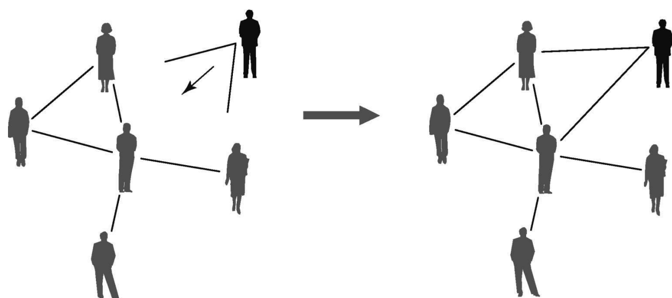


图10 在网络增长的优先连接机制中，新节点会优先连接度数高的旧节点

这一过程可从数学上加以研究，并能在计算机上进行模拟，人们可以测量真实的网络以确定其是否在起作用。一开始，所有节点多少都具备相同的度数。然而，随着节点的增长，有一些开始累积比其他节点更多的连接。某节点在特定时刻所具备的连接越多，它便越能够吸引新连接。这也是为何优先连接原则又被称为富者更富机制的原因。结果，连接性最初的小小差异会逐渐放大。因此，不同节点的层级结构便出现了，其中节点的度数差异极大，从连接最少的节点到那些累积了众多连接的枢纽不一而足。最后得到的网络具有异质性，且呈幂律度数分布。

巴拉巴西–奥尔贝特模型证明了，自下而上的增长机制能够产生异质性，而不用强加任何自上而下的设计。网络的全局标度不变性是个体及其局部选择的迭代结果：偏向连接更多的节点而不是更少的。这个模型利用概率以使个体在一定程度上可以偏离这种选择行为：一些节点可以决定连接度数低的节点。然而，总体的趋势决定结果。作为进一步的确认，我们可以检查在没有优先连接原则下成长起来的网络是否未出现异质性。的确，新节点会随机地连接到旧节点上，如此一来旧节点的度数并不影响自身吸引新连接的能力。结果出现了同质网络，其中每个节点都有着大致相同的度数。

普遍的机制

优先连接特指网络中的此种机制，而同样的机制还存在于许多自然和社会现象中，这些现象未来的演化取决于它们的历史。例如，城市规模依其目前的大小随时间而改变：大城市会经历较大的扩张，小城市则变化很小。明天的股票价格通常也与今天的价格构成一定的比例关系。这种机制也被称为乘性噪声。这一过程可能会在众多网络中起作用有着各种各样的原因。在某些情况下，具备许多连接是被新节点发现的主要方式。

与许多其他网站连接的网站要比那些连接更少的网站更容易被浏览网页的人发现。高引文献中的情况也是如此。这种增强的受关注程度使得它更容易获得更多连接和引用。

在其他情况下，连接因其自身之故而具有吸引力。社会学家发现了间接择偶这一现象的证据，即人们选择自己的配偶不仅基于相关个体的个人特征，还会考虑他人意见。例如，一项研究发现，同一人在照片中被许多其他女性环绕时，要比其独照能获得处于大学年龄段的女性更高的评价。先前一系列合作伙伴的长名单可能会让一个人处于吸引更多合作伙伴的优势地位。这便是为何优先连接也被称为受欢迎具有吸引力原则的原因。

优先连接产生的一个更加有意的理由是，与枢纽节点相连能够使相关节点更容易接近许多其他节点。自1978年放松管制以来，许多航空公司纷纷采取了枢纽政策，此政策在于选择具备良好连接条件的机场作为它们最偏爱的目的地。动机很明确：由于这些机场与大量目的地相连，与之关联便会吸引更多的潜在客户。类似的事情也发生在连锁董事网络中：任职多个董事会的人可以获取大量信息，并具有广阔视野，这又使他具备很大的吸引力从而被雇用到更多的董事会中。访问在互联网中也至关重要。互联网的大部分由私人企业建立和维护，它们被称为互联网服务提供商。当它们中有一个建立了新的基础设施，其优先考虑之事便是允许用户快速访问网络中储存的信息。考虑到这一点，互联网服务提供商并不会随机选择它们想要连接的路由器，更不会只选择更近的。相反，它们会选择那些能够保证以最少的可能步骤连接到最多服务器的路由器。那么，又有什么比枢纽节点更能实现这一点？在互联网的情况中，绘图计划所提供的数据似乎符合优先连接的假设。数字显示，某一特定时刻发布的地图中带有很多连接的顶点，往往在下一幅图中会有更多连接。

在某些情况下，优先连接会以其他机制的面目出现。设想某人想要创建个人网页。常见的办法是查看朋友的页面，挑选一个不错的用作模板。由于人们常与其朋友有着共同的兴趣，因而新网页很可能会保留模板的大多数链接，或许会改变其中少数。所以，新的页面最终将是模板的副本，变化很少。现在，这种机制掩盖了优先连接的形式。事实上，模板网页会指向什么页面呢？它们很可能会指向枢纽页面。仅仅因为枢纽页面存在并俘获了万维网中的大部分连接，任一给定页面都更可能指向枢纽处，而非那些连接较少的页面。因此，如果某个页面被复制，则每个副本都会为枢纽带来更多的连接，从而形成有效的富者更富过程，优先连接规则因而得到恢复。复制机制看起来奇怪，但实际上它在某些情况下是主导因素。例如，一篇文章中的科

学引文通常来自同一领域的其他文章，并往往会巩固该引用文章的权威性。

但复制机制最有趣的一种应用体现在基因网络中。基因组常常通过复制和多样化的过程演化。细胞复制时，所有的DNA都被复制到新细胞中，但有时候也会出错：原始DNA链条的一个完整基因被复制，并在子细胞的基因组中出现两次（复制）。大多数时候，这个新基因只是合成多余的蛋白质，这些多余蛋白质与另外那个相同基因所合成的蛋白质功能相同。然而，在进一步的复制中，两个相同基因中的一个可能会发生突变，从而会使自身合成的蛋白质行使新的生物功能，比如，与之前和自己一样的蛋白质之外的蛋白质发生反应（多样化）。许多情况中都已经发现了这种演化机制。现在，这种机制在蛋白质相互作用网络中的转化完全类似于复制机制：新的节点进入网络（复制和突变的蛋白质），它带来了一些前代的连接（即它最初与之反应的蛋白质），以及由突变而来的一些新蛋白质。高度连接的蛋白质在这种机制中占据天然优势：并非它们更可能被复制，而是它们比那些连接较弱的蛋白质更可能与复制蛋白质相关联，因而更有可能获得新连接。

尽管基因复制的作用仅在蛋白质相互作用的网络中显现，但代谢网络中也存在着支持优先连接的证据（直接或伪装）。这种关联过程的一个明显后果是，枢纽节点通常在网络里的旧节点中产生，因为它们之前有机会从“先发优势”中获益。现在，代谢枢纽正是可能在生命最初演化过程中并入基因组的原始分子，以及RNA（核糖核酸）世界的残余物，比如辅酶A、NAD（烟酰胺腺嘌呤二核苷酸）和GTP（鸟苷三磷酸），或者最原始代谢途径的元素，比如糖酵解和三羧酸循环。在蛋白质相互作用网络的背景下，人们通过跨基因组的比较发现，平均来说，演化上较老的蛋白质与其他蛋白质之间的联系要比较新的蛋白质多。

优先连接并非网络中唯一起作用的机制，也并非所有的异质网络都起源于此。然而，巴拉巴西–奥尔贝特模型还是给出了重要的关键信息。某种简单的局部行为经过多次交互迭代作用，便能产生复杂的结构。这种结构是在没有任何整体设计下产生的，即便我们允许在行为中出现一定程度的随机性，它也会出现，这解释了就整体趋势而言的个体偏差。

适应度起作用之时

当偶然的性关系在紧要关头时，人们往往对潜在性伙伴的某些特

征（例如政治观念、社会阶层或是否抽烟等）非常宽容。然而，当考虑订婚或结婚时，这些因素就变得十分重要。这是社会学家爱德华·O.劳曼对90年代中期的数据进行分析之后得到的信息。这些数据表明，美国约有3/4的已婚夫妇都具备大量的相似特征。其中包括，属于同样的社会阶层、种族群体，教育水平相同，甚至具有相近的魅力、政治观点和健康行为，比如饮食和吸烟习惯等。另一方面，当其他类型的性关系出现问题之时，相应的比例会低许多，尽管依旧很高（超过50%）。

同配生殖（即相似个体彼此结合的倾向）甚至在那些婚姻非包办、理论上任何人都能与其他任何人结合的社会中都是很强的趋势。在躁动的大学时代，个人的声望——例如，对此人之前的性伴侣数量进行衡量得出——可以成为塑造性关系网络的相关驱动力。但是，人们安定下来后，更严格的标准便会发生效力。尽管富者更富能在第一种情况中发挥作用，但它几乎无法解释第二种情况。同配生殖是同质相吸的一个具体例子：大量社会学证据表明，它是由相似个体相互关联所构成的一般趋势，也是塑造社交网络的强大力量。在巴拉巴西–奥尔贝特模型中，连接到一个节点的主要标准便是该节点的连接数，但在许多情况下，其他特征对吸引新连接而言重要得多，不管实际的连接数为多少。

巴拉巴西–奥尔贝特动力机制的一个结果是，旧节点相对于新节点而言具备累积优势。然而，实际情况并非总是如此。举个例子，万维网的旧日荣光，比如麦哲伦或Excite等搜索引擎大多已被遗忘。谷歌或雅虎等后来者已取代了前者的位置。累积优势在新的参与者进入游戏时可被完全推翻（想想脸书的情况）。新手往往具备一些固有特征，这让它们比之前的参与者更具吸引力。在这种情况下，网络的连接性并非完全由节点度数驱动，正如巴拉巴西–奥尔贝特模型中所展现的一样。相反，每个节点的某种特殊的特性能够在节点获得连接的能力中发挥十分重要的作用。这种特性被称为节点的适应度，或称为节点的隐变量，它能够塑造网络的结构，却不像连接数那般显而易见。

2002年，物理学家圭多·卡尔达雷利、安德里亚·卡波奇、保罗·德·罗斯·里奥斯以及米格尔·安琪儿·穆诺兹共同设计了一个仅凭节点适应度便可产生网络的模型。其基本思路与随机图相同：在一组节点内，考虑所有可能的配对，再依据某一给定概率看配对节点之间是否能够连接。然而，在这种情况下，概率并不固定，而是依节点在配对中的适应度而改变。模型首先会为节点分配适应度。比如，这个隐变量可能代表了个人的收入。在此情况下，适应度在节点间的分配便会类似

于国家的财富分配：先是少数十分富有的节点，然后是一定数量的上层中产阶级节点，之后便是下层中产阶级节点，诸如此类。定义连接的概率是第二步。若要模拟高度隔离的社会，则可建立以下规则：两个人形成社会关系的概率取决于他们收入的差异；具体而言，这种概率与收入差距成反比。也就是说，收入差异越大，两人获得连接的概率便越小。两个收入差别很大的人之间也可能建立联系，但这绝不是社会的主流趋势：一般而言，同质相吸原则将最终胜出。

适应度模型这种机制可能看起来过于简化，但在某些情况下，它能完美地发挥作用：例如，在世界贸易网络中便是如此。在这种情况下，适应度便是世界各国的国内生产总值（GDP）。连接的规则如下：两个国家的适应度越高，它们彼此关联的概率就越大——这将是某种适者更富的机制，其中那些GDP极高的国家会在彼此之间建立更多的商业联系。这种机制与同质相吸原则并不相同，因为，尽管GDP高的国家之间会建立许多联系，但GDP较低的国家并不会与其他低收入国家建立联系：它们的联系仍旧少得可怜。物理学家迭戈·加拉斯凯利和马里耶拉·罗弗雷多于2004年证明，这些组成部分足以精确预测世界贸易网络的特征，只要人们在模型中写入任意时刻的已有商业联系总数。例如，该模型精确地预测到了网络的节点度数分布态势。这表明，隐含在世界贸易网络这种自组织之下的基本机制已被适应度模型这种简单的动力机制捕捉到了。

跟优先连接一样，适应度机制不大可能在所有的现实世界网络中起作用。巴拉巴西–奥尔贝特模型被用于扩张中的网络时还算可靠，适应度模型对静态网络也有效，后者的节点数大致是固定的。然而，这两种模型还能同时起作用：2001年，物理学家吉内斯特拉·比安科尼和巴拉巴西共同将适应度这一概念引入到优先连接模型中，证明了这两种效应的共同作用良好预测了网络的拓扑属性。但必须注意的是，适应度模型并不总呈幂律度数分布。大范围的适应度和连接规则才能产生幂律度数分布，但很多其他的则无法做到这一点。然而，这并非某种局限性，而是这种模型积极的一面，这使得它能应用于像世界贸易网络这样的非异质网络。

各种策略

邻家女孩（或男孩）的神话已成为历史。根据社会学家米歇尔·博宗和弗朗索瓦·埃朗1989年的一项研究，80年代中期，在自己小区找到配偶的人已微乎其微了，在法国这一比例仅为3%。然而，这种情况仅在30年前还十分普遍：博宗和埃朗发现，这一比例在1914年到1960年间为15%-20%。世界上许多国家中的情况依旧如此。在某些

情况下，婚姻（以及一般意义上的社会关系）既不受某种流行趋势驱动，也不受相似性标准驱动：倘若地理限制影响很大（比如，无法定期搭乘远程交通），人们便会被迫与邻里以及同乡交友。

在这些情况下，网络的顶点被嵌入到物理空间中，这会带来许多重要后果。有时，人们几乎可以不费任何代价地与任何人建立联系（比如在虚拟网络上结交朋友）。但在其他情况下，远程联系则代价高昂。这两种情况下的网络属性也十分不同。许多基础设施网络（火车、煤气管、高速公路等）会显示出一种偏差，因为它们嵌入的是物理空间。其他网络则嵌入到时间之中：例如，科学论文会在某天发表，这会在网络连接中造成某种偏差，即新论文只能引用旧论文，而旧论文却无法引用新论文。

其他偏差和策略能够影响网络的形成。社会学家确定了人们在社交网络中建立联系的两种基本激励机制：一为基于机遇的前因，即两个人会建立联系的可能性，二为基于利益的前因，即某种促使关系形成的效用最大化或不适感最小化机制。某种数量上的全局最优化对于塑造技术网络能够发挥重要作用：例如，使万维网搜索成本最小化的压力会导致人们倾向于对最短路径长度和连接密度进行优化。

最后，甚至还可能出现这种情况：表面上的自组织起源于完全的随机状态。设想某公司发布了新的社交网络，并为10万人提供了昵称。然后，该公司许可1个人与其他1 000个人建立联系，2个人可与其他500人建立联系，3个人可与333人建立联系，4个人可与其他250人建立联系，以此类推。人们并不认识昵称背后的人，所以，他们会随机地选择伙伴。显然，这一过程中并不存在自组织：该公司建立的规则决定了网络的结构。然而，最后形成的网络在结构上仍呈现出幂律度数分布。这个例子表明，在某些情况下，幂律并不总是意味着自组织过程。

若人们试图通过为其行为建模的方式来理解网络的特征，则策略、偏差、过程和动机等一系列广泛的因素都必须被考虑进去。甚至，每个个体网络都需要自己的模型。然而，一些十分普遍的机制，比如优先连接或与适应度相关的动力机制，都可能会在许多明显不相关的众多网络的形成过程中发挥作用。本章描述的模型简单地解释了，在缺乏全局规划的情况下，局部机制何以确实能产生规模庞大、复杂、有序及有效的结构。

第七章 深入挖掘网络

谁是你的朋友？

根据上世纪90年代开展的一些研究，每有一个患有性传播疾病的美国白人，在美国某些地区就有多达20名同病相怜的非裔美国人。持续的种族不平等导致了这一结果。然而，产生如此巨大差异的真正传染机制在一定程度上仍晦暗不明。1999年，社会学家爱德华·O.劳曼和尤思科尤姆发现了一个有趣的证据：性活跃程度较低的非裔美国人（过去一年仅有1位性伴侣者）与性活跃程度更高的非裔美国人（过去一年中有4位或更多性伴侣的人）发生关系的可能性是相同情况下白人的5倍。换句话说，在白人的性关系网络中，不那么活跃的外围群体某种程度上与活跃的核心群体彼此隔离。相反，这两个群体在非裔美国人中的关联则更多。这种差异的原因尚不清楚，但其结果却很明确：在第一个网络中，性传播疾病主要在核心群体内部蔓延，而在非裔美国人中，这些疾病也溢出至外围人群。

在本例中，网络节点的度数对于理解这个现象并非最相关的变量。性伴侣数量相同的个体受感染的可能性也不尽相同，这取决于该个体是白人还是非裔美国人。在这样的情况下，仅仅知道你有多少“朋友”（即你所在节点的度数）还不够，还必须知道你的朋友有多少朋友。度数分布为图的大体结构提供了大量信息，比如它是否包含枢纽节点等。然而，度数分布并不能显示图的所有信息。比如，设想两个图具有相同数量的节点和边数：其中的节点可能有着完全相同的度数，但边的分布却可能导致这两幅图完全不同。度数乃顶点的局部特征。想要更加细致地认识网络结构，人们必须深入挖掘，并找到方法来描述节点的周边情况：距其最近的邻点，其邻点的邻点，等等。

在白人的性关系网络中，低度数节点往往与低度数节点相连，高度数节点则与高度数节点相连。这种现象又名相称混合：它是同质相吸的一种特殊形式，其中连接数类似的节点往往会互相连接。相反，在非裔美国人的性关系网络中，高度数节点和低度数节点则更容易彼此连接。这被称为不相称混合。这两种情况都显示了相邻节点在度数上的某种相关性。当相邻节点的度数呈正相关，则为相称混合；反之则为不相称混合。

通常，这些混合模式的存在是网络中某个重要机制作用的结果，这个重要机制可能就是自组织。在随机图中，给定节点的邻点完全是随机选择的：结果，相邻节点的度数之间并没有明确的相关性（尽管图的有限大小可在某种程度上掩饰这一点）。与此相反，大多数真实网络中都存在节点相关性。尽管不存在一般规则，但大多数自然和技术网络往往为不相称混合模式，而社交网络则为相称混合模式。例如，高度连接的网页、自主系统、物种或代谢物常常与其所在网络中连接较少的节点相互关联。另一方面，公司董事长、电影演员和科学文献作者往往与那些连接性与自己类似的人相关联：个体的节点度数越高，其网络邻居的度数也越高。

度数的相称和不相称仅为让节点相互关联发生偏差的大量可能相关性中的一个例子。例如，劳曼和尤姆也证明了，相比于其他群体，有更多的非裔美国人倾向于从自己的社区选择伴侣。因此，当感染进入社区之后，它便被“困”在里面了。单单这种简单的效应便让非裔美国人感染性病的可能性高出美国白人1.3倍。在这种情况下，相关性并非源于节点度数，而是与每个节点内在特性相关的一种同质相吸，这个内在特性即种族身份。另一个例子是体重的相关性：研究发现，相对于与自己体重指数不同的人而言，体重指数相似的人倾向于更频繁地在彼此之间建立社交联系。要注意的是，相关性并不必然就是支持同质相吸的正面因素：比如，在食物网中，边将植物与食草动物、食草动物与食肉动物相连，但绝少将食草动物与食草动物或者将植物与植物相连。

谁是你朋友的朋友？

科西莫·德·美第奇于15世纪带领家族接管佛罗伦萨，人们称他为“难以理解的斯芬克斯”。尽管他绝少公开发表言论，并且也从未公开采取任何形式的行动，但他依旧能够在自己周围建立起强大的党羽，并让自己成为文艺复兴时期最重要城市的国父（*pater patriae*）。1993年，社会学家约翰·F.帕吉特和克里斯托弗·K.安赛尔分析了美第奇家族与佛罗伦萨其他权势家族之间的婚姻关系、经济联系和赞助往来。他们发现，科西莫的家族位于众多权贵家族关系网络的中心。更重要者，若无美第奇家族搭线，其他家族多数时候的联系并不多，甚至彼此抵牾。科西莫的克制态度帮助自己建立起了与各家族的联盟和共治关系。

以美第奇家族为中心的网络便是自我中心网络的一个实例，在这种网络中，一组节点与中心节点（自我节点）直接连接，这组节点彼此之间也相互连接。每当后一种连接丢失一个（也就是自我节点的两

个邻点彼此不再相邻），该网络就会出现结构洞。科西莫的网络布满了结构洞，其家族能够用它们实行分而治之（*divide et impera*）策略：美第奇家族被视为许多冲突的第三方，那些家族不得不要求美第奇家族调节他们彼此的关系。

然而，对个人而言，周围布满许多结构洞并不总是件好事。根据2004年的一项研究，朋友之间不构成朋友关系的青春期女孩，其自杀概率为相反情况的两倍。这个发现的可能解释是，当事人会暴露在无关朋友的冲突之中。另外一个例子来自工会：若工人之间的联系网不存在结构洞（即自我节点被大量相互关联的节点包围时），则会形成一个强大、协调良好、交往密切的组织。一般而言，结构洞的不同模式表示不同的情况。例如，专业领域的科学家常常与该领域其他科学家相互联系，后者可能彼此也有联系。另一方面，高度跨学科领域的科学家则很可能与不同领域的科学家都有联系，后者并不必然彼此关联。

在所有这些情况下，你有多少朋友（即你的节点度数），或者他们是谁（比如他们的节点度数与你的相似还是不同）都不重要。重要之事在于，你朋友的朋友是谁：特别是，你的朋友彼此之间是否也是朋友关系。这个概念通常被称为传递性或集聚性。让我们考虑有着两位朋友的一个人：他们三人构成了连接三元组。如果此人的两位朋友彼此也是朋友，那么他们三人同样构成了可传递三元组，或者三角形。网络中的三角形数量与其中的连接三元组总数的比值便是该网络集聚系数的基本组成：这个系数衡量了该图中的三角形密度及其总体的传递性。而在随机网络中，某节点最近邻点之间的连接与任意其他两个节点间的连接具有同样的随机性。因此，这些图仅有纯粹随机连接的边所组成的三角形。另一方面，几乎所有现实世界网络的集聚系数都高于其相应的随机网络。这意味着某种重要的过程——可能是某种形式的自组织——在产生这种额外的传递性过程中起到了作用。

许多网络的高度集聚性表明存在着其中“每人都是其他每个人的朋友”这样的群体。乍一看，这幅图景似乎与网络的小世界属性相矛盾：网络是否是个“开放”世界，其中每个人之间都仅隔几步之遥？或者说网络就是紧密编织的分离群体的总和？现实中，这两个特性之间并不存在真正的矛盾：通过仔细观察沃茨–斯托加茨模型就能看出这一点。这个模型以一圈节点为开端，每个节点都与其最近和次近的邻点相连，就像遥远的村庄与其邻村交换货物一样。这是个完全集聚的结构，其中，任一村庄的所有商业伙伴也互为商业伙伴。然后，该模型为随机选择的节点开放一些连接：少数村庄开放了前往其他遥远村庄的路径，并将货物带往该处，且拒绝与其邻居做生意。少量路径便

足以陡然降低任意两个村庄之间的距离，但另一方面，我们可以认为当地紧密的商业结构已遭到破坏，也即，当地的集聚性下降了。然而，沃茨和斯托加茨发现，集聚性的降低并不像平均距离的减少那般明显。实际上，为了使传递性显著下降，人们必须重连几乎所有节点。此时，网络中仅剩随机连接了，而我们并不会期待随机图中存在高集聚性。此处的关键信息在于，网络（既不是有序网格，也非随机图）可以同时具备高集聚性和较短的平均节点距离这两个特征。

集聚性的另一个有趣之处在于，几乎所有网络中某节点的集聚性都取决于该节点的度数。通常，节点度数越大，集聚系数便越小。低度数节点往往属于彼此连接良好的局部网络。类似地，枢纽节点与众多节点连接，这些节点并不会直接地彼此连接。例如，互联网中低度数的自治网络常常属于高度集聚的区域网络，且经由全国主干网彼此连接。许多网络中都可能出现类似的结构，其中集聚性会随着节点度数的增加而呈下降趋势。

谁是你朋友的朋友的朋友……？

金钱会在一定程度上带来幸福，但周围快乐的人却能给人们带来更多的幸福。根据1984年的一项估计，每年多挣5 000美元会增加2%的幸福感，而社会学家尼古拉斯·克里斯塔基斯和詹姆斯·福勒2008年的一项研究表明，拥有一个快乐的朋友则会提升人们15%的幸福感。两位社会学家调查了来自马萨诸塞州弗雷明汉的超过12 000名居民的主观幸福感。不仅如此，他们还绘制了这些人之间的朋友、配偶或兄弟姐妹关系。通过绘制这个关系网，二人发现，联系紧密的人常常有着类似的感觉：幸福的人往往聚在一起，另一方面，不幸之人亦是如此。克里斯塔基斯和福勒甚至还发现了更多有趣的证据。个人的幸福会受到其非直接相邻的人的幸福程度的影响。两步之外（朋友的朋友）的“幸福效应”约为10%；三步之外（朋友的朋友的朋友）则为约6%。这种效应仅在第四步就消失了。这两位社会学家和其他科学家也在肥胖、吸烟习惯以及口头建议（比如找一个好的钢琴教师或一个好的宠物之家）等方面发现了类似的结果：在所有这些情况下，三度空间之外的影响或信息会对个人起作用。人们在不同社会过程中发现的这个三度空间规则是超二元扩散的一个例子，即超越连接最近邻居的二元关系的扩散现象。在这种情况下，每个节点的度数、其邻点的度数，还有邻点之间的连接都不再重要。影响超越了各节点最近的连接圈。实际上，在许多现象中，这种影响甚至超过了三度空间。例如，高度传染性疾病可形成更长的传染链；类似地，营养物则可扩散至整个食物网。

在这种动力机制中，节点的重要程度取决于通过它的链条数量。为了捕捉这个观点，社会学家林顿·C.弗里曼引入了节点的中介中心性这一概念。取某网络中的所有节点对，数一数关联它们的最短路径数。节点的中介中心性基本上就是穿过该节点之最短路径占所有路径数量的比例。这一比例越高，相关节点的中心性就越高。按照这种测量方式，美第奇家族便是15 86世纪佛罗伦萨家族中最具中心性的一个。在这种情况下，中介中心性便可衡量减缓节点流或扭曲通过链条的可能性，它以这种方式服务于中心节点的利益。一些研究表明，企业在经济网络里的中心性很好地预示了它的创新能力（根据获得的专利数衡量）以及财务业绩。有趣的是，1980年到2005年间，东亚国家在世界贸易网络里的中心性经历了大幅增长，而大多数拉美国家却呈下降趋势。然而，这两个地区的贸易统计显示出类似的模式：宏观经济统计并未很好地追踪二者发展的巨大差异，而基于网络的方法却捕捉到了这一点。根据1965年的一项研究，莫斯科在中世纪便成为俄罗斯中部河流运输网络最具中心性的节点。很可能，这为其未来的重要性打下了基础。

中心节点通常充当桥梁或瓶颈：它们几乎是网络交通中的必经站点。因此之故，中心性乃是对网络节点的负荷的估计，前提是大多数节点之间的连接都经过最短路径（情况并非总是如此，但这是个很好的近似情况）。由于同样的原因，中心节点的损坏（例如，某个中心物种灭绝或某个中央路由器被破坏等）会从根本上影响相关网络的节点连接数。根据想要研究的过程，还可以引入中心性的其他定义。例如，接近中心性计算某节点到其他所有节点之间的距离，而抵达中心性则将网络内所有节点分解为经由一步、两步、三步等不同步数抵达的节点。此外中心性还有一些更为复杂的定义。

许多现实世界网络的中心性特征是它们异质性的深层标志。许多真实世界的网络会表现出异质分布特有的长尾。平均中心性并非对任一节点的有效估计，因为这一量值在平均范围内变化很大：少数节点便是网络中几乎所有最短路径的主要瓶颈，整个中心性较低的节点层级都要经由他们向下。考虑到中心性较高节点的重要性，我们很自然地会问它们是否与网络的枢纽节点相同。很多情况下，事实的确如此。例如，高度连接的自治系统也可作为区域网络的桥梁；多义词因与许多其他语词连接，从而将语言中的不同领域联系在一起。但这并非普遍规律。机场便是一个显著的例外：其中，某些低度数机场具有特别大的中介性。2000年，中心性最高的机场为巴黎机场，它是关联着250多座城市的枢纽节点。中心性次高的机场则为位于阿拉斯加州的偏远的安克雷奇，它是一个仅与40座城市连接的中等大小的机场。其他类似机场也出现在最具中心性的机场名单上。这种异常又该如何

解释？阿拉斯加州有许多飞国内航线的机场，但安克雷奇是通往美国其他地方的唯一桥梁，因此，许多航线都穿过该机场。这种异常是局部地区机场密度高，却少有通往国外的航线的结果。

你属于什么群体？

1972年，美国一所大学俱乐部的两名空手道教练发生了激烈的冲突，从而决定将他们的俱乐部一分为二。这件在世界上多数人看来不值一提的小事却成了社会学家韦恩·W.扎卡里眼中的一座金矿。1977年，他发布了一项关于此事的开创性研究，并在其中提出了一个新奇的观点。

1970年，空手道教练希先生要求俱乐部主席约翰·A.提高课程价格，以提供更好的薪酬。而他所得到的只是拒绝。随着时间的推移，整个俱乐部都因此事而产生了分歧，两年后，希先生的支持者们在其领导之下组建了一个新的空手道组织。在此期间，扎卡里搜集了空手道课程、会议、派对以及俱乐部成员的聚会等相关信息，并将那些在俱乐部之外还见面的人定义为好友。这时，他便能为该俱乐部绘制一幅精确的友谊网络图谱了：所得图形的结构明显围绕两位教练而分为两个群体，每个群体里的人都互为好友且与其中一位教练交好，而两个群体的成员之间却很少往来（图11）。当俱乐部一分为二，人们几乎都沿区分两个群体的界线站队。

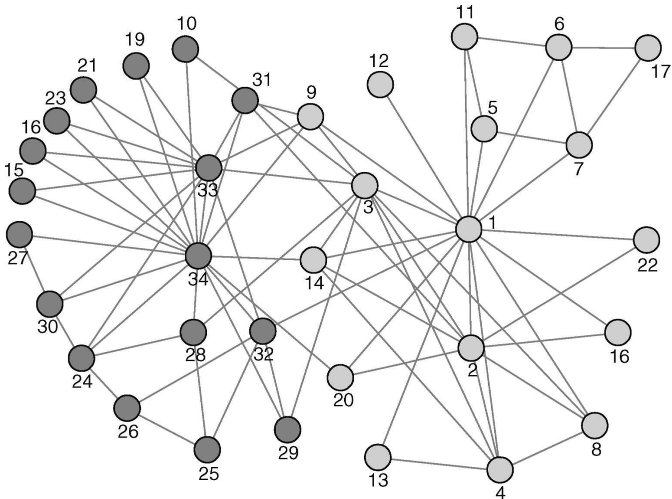


图11 人类学家韦恩· W.扎卡里研究的空手道俱乐部中的友谊结构能让我们预测到该群体会一分为二

仅仅基于网络结构本身，扎卡里的方法便能够几乎完美地预测俱乐部的分裂。从那时起，研究人员便一直致力于找出能够识别网络中的社区或模块的通用办法。在扎卡里的例子中，这些社区或模块仅通过查看图形便能看出，但在其他复杂得多89的情况中，人们尚未发现通用的解决方案。所有真实世界的网络都在一定程度上显示出模块化特征。很明显，阿拉斯加机场便是机场网络结构中一个特定的模块，其他内部连接良好却与外部没有连接的区域也是如此。食物网也分为若干不同的分部，即那些内部互动更加频繁而与其他物种联系较少的物种群体。社交网络也分为不同的团体：例如，针对青少年的研究表明，他们的行为强烈地受其所属群体的影响。神经网络常被划分为对应特定功能的大区块。基因调控网络则被分为不同的子网络，后者与特定的功能或疾病相关联。度数、相关性、集聚性以及中心性都提供了单个节点及其紧邻的周遭环境和节点在整个网络中的相对地位等信息，但它们并不体现整个图形所分解成的各别结构。

模块的更简单形式是模体，它是少数节点在整个网络中重复出现的连接模式。在食物网中，我们经常会发现某种菱形结构：例如，某种食肉动物捕食两种不同的食草动物，后两者食用同一种植物。另一种常见的模体是三个物种的简单链条：大鱼吃小鱼，小鱼吃虾米。这些模式并非纯粹概率的结果：模体在真实的食物网中出现的频率比在其随机对照网络中高出很多。通常，在大型网络中，人们可以区隔出许多可能为候选模体的由节点和边组成的子集。然而，只有当给定的子图出现在某网络中的频率高于其随机对照图时，该子图才能被认为是相关模体。在万维网中，一个十分常见的例子是二分团：它由两组网站组成，其中一组的所有网站与另一组所有网站之间相互连接。通常，这种模体能够确定一组有着相同兴趣的“粉丝”群体（比如有关漂流的博客），并指向他们的“偶像”（例如漂流杂志的网站等）。调控基因的网络则几乎完全由模体构建。当大肠杆菌处于应激状态时，特定的基因回路会感知到应激状态，并协调某些蛋白质的产生。这些蛋白质联合起来形成鞭毛，这种不断摆动的尾状物能让细菌游走以寻找更好的环境。相同的遗传回路，即协调前馈环还存在于许多其他细菌和一些其他微生物体内。演化似乎已经因为特定模体的最优特性而选择了它们（例如，因为它们能用更少量的所需基因来执行某项功能）。此外，模体机制的明显优点是模体可以组合产生新的功能，而其中某个模体受到破坏也不会影响到别的模体。

模体乃某种小规模、局部的重复性模块。但当人们考虑社区时，他们通常意在发现网络中的大型分区，比如食物网的区划、在线社区、学科领域等等。这些结构并不表现出规律性的重复模式。如果我们掌握了某些线索，比如，假如社区的成员因某个因素而自发确定自

己的身份，则相对容易找到它们。这个因素可能是社区内所有成员添加到他们博客上的小饰物、共同的着装方式等。然而，多数时候这种信息既非现成也不明确，我们必须深入挖掘网络结构以找到模块。社区识别的总体目标为发现那些内部连接比彼此之间连接更为紧密的节点集，就像空手道俱乐部网络中呈现的那样。口头表述是很容易的，但将此概念转换为数学表达却很难，以至于人们尚未发现确切的社群检测方法。一些方法可以聚合节点以满足最优性准则。其他方法则可将网络拆分为群组，然后再进一步将群组进行拆分，接着再进行拆分，进而创建一个嵌套社区的谱系树。还有一些方法在节点之间放置假想的弹簧，然后查看系统松弛之后所形成的节点群集。一般来说，还有很多其他方法上的选项。有一种有趣的技术聪明地利用了网络拓扑学，其基础在于计算边介数，也就是找出多数最短路径所通过的边。具有最高边介数的连接类似于在格兰诺维特的研究中连接原本相互分离之群体的弱连带。如果去掉一些高介数的边，那么，网络就会分裂成一定数量的孤立节点群集：它们便是恰当的候选社区。我们还可以继续去掉网络中一些介数较高的边，以找出嵌套在更大结构中的更为精细的结构。

社区发现方法的一个有趣应用是对美国政治博客圈的分析。物理学家拉达·阿达米克在民主党人和共和党人的博客间发现了清晰的分隔。其得出的网络结构显示，这两大党的阵营绝少相互关联。此外，民主党的博客比共和党的博客更缺乏凝聚力。例如，在此博客圈的堕胎讨论专区中，反对堕胎的博客比支持堕胎的博客联系更为紧密。因此，在线造势运动更可能在前者的博客中得到传播。另一项研究分析了美国学校中的学生社区特征，借此了解种族是否会塑造社交网络。在种族非常多元和十分同质的学校中，这一因素似乎都是无关紧要的。相反，在种族多样性处于中间值时则能看出隔离特征。在代谢网络中，人们已经发现了与特定功能相对应的社区（碳水化合物代谢，核苷酸与核酸代谢，蛋白质、肽和氨基酸代谢，脂类代谢，芳香族化合物代谢，单碳化合物代谢和辅酶代谢等）。最后，公司股票则在价格相关性的基础上聚类，人们能从中发现与银行、矿业、分销、金融等各业务领域相对应的模块。

将社区定义为“内部联系比外部联系更为紧密”的子图非常普遍，但这种做法并未涵盖某些特定的模块。想想朋友间的电话通信链条，其中第一个呼叫第二个，第二个呼叫第三个，以此类推：根据上述定义，这种链条极有可能不会被归类为社区。另外一个例子则是同行业竞争者的网页：显然，他们并没有动力相互连接，尽管他们明显属于同一社区。此外，真实世界的社区比密集的节点群集复杂得多。多种划分可能同时出现；国籍、社会阶层、性别、工作、政治观念统统都

可用来对同一群人进行分类。而且，社区之间可能互相重叠：一个人可能同时从属于多个国籍或隶属关系。最后，嵌套社区也可能存在：例如，地域身份从属于国籍。

尽管过于简化，但图示法仍然能够捕捉到系统的诸多相关特征。若我们仔细查看图，便会发现大量相关信息，而运行的复杂测算越多，便会呈现越多细节。真实世界的网络几乎在任何时候都会偏离其随机对照网络，这意味着其中存在某种嵌入的秩序。同样，所有这些网络都未经过设计：偏离很可能产生于自组织过程。目前，在图的结构中寻找新的规律并揭示其潜在机制是网络科学仍然面临的一些挑战。

第八章 网络中的完美风暴

意外之背景

巴罗科罗拉多岛是巴拿马运河中央的一块热带雨林。当其附近的一条河被筑坝拦截后，该岛便只剩下几处小山顶还显露在水面上。它现已成为一个露天试验场，而试验内容是高速公路、建筑物、田地或矿井替代了原始植被，雨林被分割为若干小块之后的情形。洪水淹没巴罗科罗拉多岛周边数年之后，美洲虎和美洲豹的种群数量都迅速萎缩了。结果，它们的猎物种群数量迅速攀升：现在，一种叫刺鼠的大型啮齿类动物已遍布该岛。这些啮齿动物爱好金合欢的硕大种子，所以，它们的繁盛对金合欢的成功繁殖以及依靠这些种子维生的微生物都构成了巨大难题。随着金合欢种群数量的减少，那些种子较小的植物便取而代之，而以后者为食的动物数量随之增加。生态系统的原初变化便向岛上食物网的各个方向延伸开去。

食物网中的多米诺效应并不罕见。通常，网络一般会为大规模的突发及意外动态提供背景支撑。运输系统中的病原体、电网中的断电、社会系统中的大型冲突或意想不到的合作努力等都证明了：网络似乎是“完美风暴”的理想背景。网络节点代94表交换物质或信息（信息包、能量等）的单独实体（人、计算机、物种、基因……），或者它们也可以表示交换单独实体（货物、旅行者……）的场所（国家、机场……）。在这个非常宽泛的分类中，潜在动态的范围十分广大。为何网络会成为这些动态的天然发生场所？图的结构又会如何影响这些过程？对于这些问题不可能有总体答案，但在许多情况下我们能看到，底层网络的异质结构、非随机组织等特征会对表层发生的所有现象造成重要影响。

故障和攻击

2001年7月18日，一列火车在美国巴尔的摩的地下隧道中脱轨并起火。稍后，美国东海岸几个州的网速就变慢了。大火烧毁了途经隧道的光缆，这些光缆分属于美国最重要的几个互联网服务供应商。结果，这一事故引发了横扫美国大部分地区的多米诺效应。互联网常常面临类似的事故。任何时候总会有一定比例的路由器由于各种原因而

一直处于死机状态，而每次类似事故都可能会与巴尔的摩脱轨事故一样严重。然而，这种大范围的破坏还是很少见的。网络似乎会容忍一定数量的长期功能障碍而不会出太多问题。这在一定程度上依赖于替代路径，后者允许网络内部流量绕过故障区域。然而，互联网与多数网络一样并没有太多冗余连接，其密度也并不是很高。考虑到这些特征，我们很自然就能预料到它比较容易发生故障。

尽管互联网似乎相对能够承受一些错误和意外，但精心设计的一次攻击仍能造成严重破坏。2000年2月7日，大量用户登录了雅虎网站，其规模之大令该公司的服务器无法响应这些请，网页于是随之崩溃。在随后的几天里，一系列其他网页（从趣到美国有线电视新闻网络等）也因为同样的原因而崩溃。两个月后，警察发现此类登录乃一名15岁的加拿大黑客所为，其昵称为“黑手党男孩”。他并不需要烧毁任何电缆便能阻塞互联网：其所做之事足以让这些吸引万维网上多数流量的网站崩溃。

跟互联网和万维网的情况一样，大多数真实世界的网络都显示出一种双刃的鲁棒性（robustness）。即便大部分网络遭到破坏，它们仍然能够正常运转，但某些突然的小故障或者有针对性的攻击则可能让它们彻底崩溃。例如，基因突变在整个生命过程中都会自然发生（其中有些甚至能删除细胞中的某些蛋白质），或者也能人为地产生（一项名为基因剔除的基因技术就是如此，它能关闭实验室老鼠一个完整基因的功能）。但是，生物体仍能表现出对诸多突变以及大量基因剔除的极强忍耐力。多数时候，生物体还是会在整体上继续正常工作。另一方面，某些特定的突变却能完全破坏细胞的工作。大脑一直在丢失神经元：一次给任一器官带来压力的经历，比如偶尔的酗酒，就会杀死大量细胞。但宿醉过后，一切又都恢复如初。以帕金森症为例，相当比例的神经元甚至会在病人毫无察觉的情况下消失。但当这一比例超过某一阈值，受损情况便开始变得明显。

在这方面，网络与设计而成的系统十分不同。以飞机为例，一个元件的损坏便足以让整个机器停止运转。为了让它更具复原力，我们必须采取策略，比如复制飞机的某些部件：这能让它几乎100%安全。相比之下，多数并非设计的网络则对广泛的错误表现出自然的复原力，但当特定元素失效，它们便会崩溃。网络能容纳多少错误而不出问题？而导致其崩溃的因素又是什么？为了回答这些问题，科学家通过移除网络节点以观察会发生什么情况的方式来模拟故障。删除一部分节点之后，他们会检查剩余的节点是否仍旧相互关联（即某种巨型连通分量是否仍存在于网络之中）且连接紧密（即节点之间的平均距离是否依旧很小）。为了模拟误差，节点是随机移除的。当在随机

图中如此操作时，几次移除就会导致节点间距离迅速增加，图也瓦解为许多不相连部分。在半数节点移除之后，与大多数真实世界网络大小一样的随机图便遭到破坏。另一方面，异质网络（不管是真实网络图还是大小类似的无标度模型图）中经历相同的过程时，其中的巨型连接通量在80%的节点都移除之后仍然存在，而其内部的距离则实际上与最初无异。而当研究者模拟像“黑手党男孩”使用的那样一项针对性攻击时，情况则有所不同。他们一开始移除了网络中最“重要”的节点（枢纽节点）。在这种情况下，两种网络的崩溃速度都比之前快很多。然而，后者更为脆弱：在同质网络中，需要移除大约1/5的枢纽节点才能将其摧毁，而异质网络刚被移除少数枢纽节点，就会发生坍塌。

高度连接的节点似乎在错误和攻击中都发挥着至关重要的作用。它们是暴露在针对性攻击之下的多数异质网络的致命弱点。在这些网络中，枢纽节点主要负责图的整体聚合，移除其中少数枢纽节点便足以将其摧毁。另一方面，枢纽节点也是这些网络在暴露于错误和故障中时的“王牌”：当随机移除节点时，多数时候被选出的节点均为低度数节点，因此，只要枢纽节点保持不变，网络便不会坍塌。考虑到节点度数常常与中介性相关，这种情况就会愈加清晰。高度数节点多数时候都是许多网络路径经过的桥梁。当网络遭到随机破坏，为数不多的枢纽节点很少会受到影响。既然枢纽节点不受影响，它们便提供了必要的连接：许多冗余连接则显得多余；经过枢纽节点的路径会让受损网络的工作区域保持连接状态。在某些低度数节点具备高中介性且扮演桥梁角色（就像某些机场一样）的少数网络中，枢纽节点遭到攻击仍然会导致严重的破坏，但最致命的策略还是攻击最具中心性的节点。

多米诺效应

系统可能会从能够容错的弹性状态突然转向全局性崩溃，这应该引起人们的警醒。根据估计，一定速度的物种灭绝在生态系统中是不可避免的：每年，每一百万个物种中就会有一个灭绝。通常，食物网会在灭绝事件发生之后重组，绝大多数物种并不会受到此类自然灭绝事件的重大影响。但大规模的食物网崩溃也是可能的：大约2.5亿年前，超过90%的物种在相对较短的时期里纷纷灭绝，这便是著名的二叠纪大灭绝事件。过去5亿年里，地球上总共发生过5次类似的大灭绝事件。研究人员认为地外因素可能是罪魁祸首，比如充满争议的可能导致恐龙灭绝的陨石坠落事件。然而，也可以用网络来解释这些事件。物种的连环灭绝或共同灭绝事件对于生态学家而言并不陌生。例如，英国曾于20世纪中期引入黏液瘤病毒以控制兔子的种群数量，这

最终导致大蓝蝶于1979年灭绝。这种病毒消灭了兔子，但兔子食用的高茎草却因此得以蔓延。后者又破坏了蚂蚁的栖息地，它们习惯在有阳光的低茎草中筑巢。蚂蚁与大蓝蝶的幼虫之间有着共生关系：它们会照看幼虫，而幼虫则报之以流食。因此，蚂蚁栖息地的破坏逐渐损害了大蓝蝶的繁衍，进而导致其灭绝。这并非真正意义上的共同灭绝，因为兔子并没有因为黏液瘤病毒而灭绝，而部分大蓝蝶也被重新引进。然而，这一事件却让我们对食物网可能受损的程度有所了解。将此事件的规模扩大，其中的连环灭绝几乎耗尽整个生态系统的物种，这便可能成为以往大灭绝事件的另一个解释。当人类主动对生态系统进行大规模破坏时也应考虑这种情况，例如，目前过多的捕捞正以前所未有的规模消耗着海洋生态系统。

网络上少数其他动态过程也可能引发类似的连锁故障或雪崩式崩溃。大规模停电便是一个典型例证：某个发电站的故障导致另一个发电站过载，这又导致后者出现故障，继而在网络中大范围传播电力过载故障。在此现象中，某一节点的故障不仅会导致节点互联的损失或降低节点之间的平均距离，而且还会引发多米诺效应。金融危机中出现的经济网络的系统故障则是这种现象的又一例证。拥堵现象中也会发生同样的情况，比如车流致使街道网络某些地点的交通发生瘫痪，人流因特殊事件而在地铁站中形成阻塞，又或者网络流量导致某些互联网服务器崩溃，等等。研究表明，枢纽节点在所有这些情况中都至关重要，不仅因为它们能减少运输时间，还因为它们会率先饱和。

流行病

1347年，一场人类史上最具破坏性的瘟疫在君士坦丁堡暴发。之后的三年里，黑死病蔓延至欧洲，导致该地区大部分人口死亡。这种疾病像波浪一样以每年200-400英里的速度在欧洲蔓延开去（图12左）。这种传染图景与现代流行病大异其趣。据估计导致世界3%人口死亡的1918年大流感仅用了一年时间传播，它甚至蔓延到了与世隔绝的太平洋岛屿上。1957年的“亚洲流感”病毒仅用了半年时间便席卷全球。而更近的2009年的猪流感则在几星期之内便从地球一侧蔓延到了另一侧（图12右）。黑死病潜伏在船只和车厢中经由朝圣者、商人和水手等媒介传播的速度每天不过数英里，而现代疾病则依赖更高效的交通方式传播，比如高速公路、火车和飞机等。14世纪时，物理距离是流行病传播的主导因素。而在现代的网络世界中，传染病能跳上飞机，数小时之内便可到达地球的另一端。

流行病经由全球网络（比如机场网络）和地方网络传播：人际传染病则依靠个人社交网络传播。例如，流感在一定程度上通过个人之

间的面对面接触传播，艾滋病病毒则经由无保护性接触网络传播。2001年，卡比兰公司的物理学家罗慕阿尔多·帕斯托尔·萨托拉斯及其意大利同事亚历山德罗·韦斯皮尼亚尼决定通过建模和模拟疾病在社交网络中的传播来研究这个问题。他们只引入了疾病传染过程的最简机制：一开始，社交网络中的少数个体感染了疾病；如果某健康个体通过某种关联与这些人中之一有过接触，则他或她有一定的概率会被感染；另一方面，被感染个体也有一定的概率会康复。这种感染模型被称为“SIS”因为每个人都会经历“易感”，–感染–易感（susceptible-infected-susceptible）这个周期（健康个体对疾病“易感”）。这个过程表现了类似普通感冒的传染病，感染人群往往会在这个过程中恢复。还可以将这个过程复杂化以表现其他传染病，例如引入致死或免疫等可能性。然而，这些修改无法改变最终结果的大致走向。

14世纪的瘟疫像波浪一样在欧洲蔓延（左），而2009年的猪流感则更像火团从地球的一端至另一端溅下火花（右）其中的差异在于人类交通网络的巨大变化

萨托拉斯和韦斯皮尼亚尼发现，病毒在最初的扩张之后要么被根除——迅速减少并最终从人群中消失——要么成为地方病——停留在某地并反复感染该地区的部分人群。若每个被感染个体所感染的人数少于一人，则该种疾病低于传播阈值：在这种情况下，该疾病会逐渐消失。若每个被感染个体所传染的人数超过一人，则该疾病便已超过传播阈值：在这种情况下，该疾病会逐渐传播开去。若疫苗可用，人们便可通过使足够比例的人口免疫而将相关疾病控制在传播阈值之下。高传染性疾病常常最为棘手，因为它们的传播阈值很低，进而十分容易传染开去。如果人们难以根除某种疾病，则将其传染性降至传播阈值附近仍有积极效果：这会降低受地方性疾病反复感染的人口比例。

在萨托拉斯和韦斯皮尼亚尼的研究中，他们发现传播阈值主要取决于底层网络的特征。当SIS机制作用于随机网络时，出现了一种能够使我们估测根除相关疾病所需免疫人数的明确阈值。但当这种机制作用于异质网络时，免疫阈值则几乎消失：它远低于随机图的阈值；此外，系统规模越大，免疫阈值越低。若网络足够大，免疫阈值会低至几乎无法避免部分人群被感染的程度。不将几乎全部人群进行免疫是无法将疾病控制在如此之低的阈值下的。流行病与其他许多动力机

制一样，都会受网络异质性的影响。人们对故障和攻击的研究已经表明，枢纽节点会将网络的不同部分连接。这意味着它们也扮演着传播疾病之桥梁的角色。枢纽节点的众多联系将它们与被感个体和健康个体相互联系：因此，枢纽节点很容易被感染，它们也容易感染其他节点。流行病学家确认的超级传播者很可能就是社交网络的枢纽节点。

异质网络在流行病面前十分脆弱不是个好消息，但加深对它的理解却会为疾病控制提供好的思路。理想情况下，几乎所有群体都应被免疫以完全阻断传染。然而，如果我们只能免疫一小部分人群，随机选择免疫人群便不是个好主意。多数时候，随机选择意味着会选出那些与他人联系相对较少的个体。即便这样能让他们阻断疾病在其周围的传播，但枢纽人物还是会让疾病重新传播开去。瞄准枢纽人物是更好的策略。对枢纽人物进行免疫就像将其从网络中删除，而对针对性攻击的研究表明，删除小部分枢纽节点会打碎网络：因此，疾病将被限于网络中少数孤立区域。这个策略面临着一个实际问题：无人真正知晓某个人群的完整社交联络图，所以很难确定枢纽人物。然而，物理学家鲁文·科恩、什洛莫·哈夫林以及丹尼尔·本–亚伯拉罕在2003年提出了一个精巧的策略以找到枢纽人物：他们建议对人群随机抽样，并询问与这些被选取人相互联系的人的名字。这个名单中重复最多的名字最可能是社交网络的枢纽：事实上，枢纽人物因自身的众多连接而与很多人相关联，所以许多受访者都会提到他们。我们应当注意，免疫枢纽人物在理论上十分奏效，但真实世界的诸多细节会降低其有效性，比如，网络是否会处理枢纽节点周围的特殊冗余路径，或者联系网是稳定不变还是在不断演化等：例如，若爱丽丝携带艾滋病病毒且与鲍勃有过无保护性行为，而鲍勃又与卡罗尔有过无保护性行为，这两次性行为发生的先后顺序对卡罗尔而言则意义重大。

流行病在社交网络中的传播图景可部分推广至其中节点不表示人群而代表地点（比如机场）的网络中，并且在这个网络上传播的是人群（比如被感染或健康的旅行者）。在这种情况下，我们可以定义一个全球入侵阈值，高于此阈值的疾病成为流行病，反之则为局部层面的疾病。关闭机场往往并非上策：我们需要关闭90%的机场才能有效阻止某些流行病的传播，这将带来巨大的社会和经济损失。与发展中国家（这些国家通常是新型流行病的源头）共享抗病毒药物这种更聪明的策略则有效得多。

病毒、广告与时尚

两个无名的巴基斯坦程序员，一位大学教授，一群高中生……这些人便是计算机病毒的始作俑者。上世纪80年代，这些寄生病毒程序

开始在计算机之间传播，它们基本上藏身于用户互换的软盘上。第一批病毒是自我复制软件的学术实验品，但它们很快从实验室中逃逸。1986年，“大脑”病毒出现在巴基斯坦。同年，一家德国实验室失去了对“波光”病毒的控制。一年后，一群学生传播了“维也纳”病毒。90年代，计算机病毒已成为全球性问题，但这与即将发生的事情相比则微不足道。

互联网的出现也带来了新一代病毒，它们能通过网络将自己拷贝发送至其他计算机。1999年，“梅丽莎”病毒在互联网蔓延：人们开始收到主题为“有你的重要消息”或“这是你请求的文件……请勿转给他人；”-）的邮件信息。这些邮件中包含一个名为“list.doc”的文件。如果接收者将其打开，它就会启动一个程序，进而将相同的信息发送至计算机中保存的前50位联系人。“爱虫”“蓝宝石”“巨无霸”“冲击波”以及其他许多计算机病毒纷纷在网络上爆发，它们有着类似的机制并都造成了灾难性后果：其中一些破坏了公司的计算机系统、大学的数据库，甚至影响了互联网流量。

计算机病毒传播中的一些特征与真实世界的流行病惊人地相似。计算机经由自己的连接（例如，作为计算机主人社交联系人样本的电邮网络）而受到感染，同时也以类似的方式感染其他计算机。一些关于疾病的结论解释了计算机病毒令人困惑的行为。即便杀毒程序会及时更新，一些病毒仍会在其首次攻击后的数年里继续传播。若考虑到流行病在无标度网络中的流行特征，这便不足为奇了：即使大部分计算机因杀毒程序而免疫，也不足以根除病毒感染：总会在这里或那里存在一些高度数节点让病毒重新流行。

若想在异质网络中传播信息，对计算机病毒构成严重问题的异质网络容错特性则可转变为某种资源。这便是病毒式营销背后的原理。多亏了虚拟社交网络，万维网如今充满了“像病毒般扩散”的视频、游戏和应用程序：每天都会有成千上万人将这些内容转发给自己的所有联系人。体现这个想法的首个例子是微软电邮服务的传播。1996年，微软公司在邮件中插入写有“来微软电邮获取你的免费网页邮箱”的自动页脚，其中包含的链接能让人在很短的时间里设置一个新邮件地址（见本书第50页）。类似的策略也体现在雅虎、谷歌的邮箱服务以及许多基于邀请才能使用的社交网络服务中。

病毒式营销利用了一种被称为社交传播的潜在心理现象。这是人们模仿其联系人传播闲话、时尚、谣言和观念的大体趋势。这种心理机制也会在创新、团队解决问题和集体决策中起作用。社会学家和心理学家们发现了人类带有明显彼此“模仿”倾向的很多例子。一一1962年，坦桑尼亚所教会学校中的群女孩就经历了一场无法控制笑声的非

正常倾向。数月之后，同一所学校里的几十名学生也都表现出相同的症状，而这些学生中的一部分被送往一些村庄安置后，村中其他人也表现出同样令人不安的咯咯笑的症状。经过大量调查，研究该病例的医生A.M.兰金和P.J.菲利普得出结论：这是“集体歇斯底里”的一个病例。1998年田纳西州的一所高中发现了一个类似案例，当时，感觉闻到汽油的老师将这种感受传递给了数百名学生。排除所有外部环境因素后，科学家们得出结论说，某种“情绪传染”机制在其中起了作用。

人们已记录了许多类似的社交传播案例，但近年来科学家们发现，同样的机制还可能在不那么特殊的环境中起作用：例如，肥胖和吸烟似乎也会在社交网络中传播。相互关联的人群共有某种特征或行为主要基于三个原因。首先在于他们属于同一社会阶层这种外部因素：例如，同属较低社会阶层的人吸烟和肥胖的风险较高；同时，与社会阶层较高的人相比，他们更有可能与彼此建立联系。其次，人以群分：吸烟者或那些有着相似体重的人们往往会与有着类似癖好的人成为朋友。第三，社交传播的作用：如果你是吸烟者或超重者的朋友，那你更有可能开始吸烟或增加自己的日常食物摄入量。这三种因素可能同时起作用，但社交传播可能最为重要，不应被低估。社会学家认为，传播的并非某种特定的情况；相反，是人们对何谓恰当的共识得到了传播。这种视角可应用于公共卫生领域，通过采取针对社交网络枢纽节点的措施来培养人们更加健康的习惯。

自然，行为、谣言及观念的传染在许多方面都与疾病传播有所不同。与传染病不同，散布信息必然是有意为之。另一方面，获取信息往往对人有利，因此它更多地是一种主动而非被动感染的过程。学习或被说服所需的接触时间可能比患上疾病要长。此外，许多其他竞争因素也会起作用。如果社交传播是主导因素，均一性将是其内在规则，但事实上，反对简单同化的因素会产生多样性、少数群体和两极分化。无论如何，社交传播在一定背景下或许的确是最为相关的因素。理查德·费曼于1940年代发明了费曼图这种现代高能物理学工具。一些物理学家满心热诚地接受了这种工具，另一些人则对其抱有疑虑，但这些图最终成功了。针对费曼图在美国、日本、苏联物理学家团体中扩散的研究表明，其扩散趋势可与流行病模型十分准确地吻合，前提是将模型中的参数调整为十分不同的值。

网络及动力机制孰先孰后？

古罗马成功的关键因素之一是其紧邻台伯河的战略地位，该河当时乃首屈一指的通信和商业路线。城邦变得更加强大之后，人们便开始营建其强大的道路网络分支。反过来，道路网又成为维持并进一步

扩张罗马强权的关键工具，因为路网使得运输货物和军团更加快捷。更多的道路意味着更多的权力，而更多的权力又必然会打造更多道路：结果便如意大利谚语所言，“条条大路通罗马”。我们几乎能在每个重要城市的历史上发现类似的发展模式。扩张中的城市吸引了更多的交通，也需要更多的连接方式（道路、铁路、航线……），这些因素反过来又会增加交通流量和扩大城市规模，而这又意味着更多的连接。通信网络会影响交通的动力，后者反过来又会在反馈回路中重新塑造网络。

网络拓扑如何影响了动力机制这一问题暗含了某种假设：网络乃不可变的结构，过程在它上面发生。现实中，所有网络都会在动力作用的过程中发生改变。因此，仅当动力机制的发生时间比拓扑早很多时，这一假设才有意义。在某些过程中这是合理的：例如，人们每天或每周相互交换的信息会在固定的社交网络中传播，因为通常友谊与亲属关系的更替是按年进行的；或者，城市的车辆交通不论在哪一天都会在固定街道上进行：通常街道连接不会每天都改变。

然而，在其他情况下，这种假设是有缺陷的。例如，在性传播疾病的传播过程中，与他人发生关系的时间顺序尤为重要。与某人建立无保护性关系发生在与受感染的他人建立无保护性关系之前还是之后很不一样。如果我们想研究一个城市十年的发展，那么有必要考虑交通和正在改变的连接之间的相互作用。在类似对等文件共享系统等一些技术网络中，网络结构和信息动力机制在相同的时间段内变化且强烈地相互交织。而食物网中的种群动态能够引起网络的重组。当过度捕捞将某物种数量降至一定水平之下时，食物网会重新排列捕食次序，并让新物种代替旧物种。网络结构和动力机制的耦合在虚拟社交网络发展的特定时刻尤其重要。这些工具提供了个人网络结构和内容的恒定信息流。因此，研究者认为，这种增强的意识可能改变人们创作、维护和影响其社交网络的方式。

一些方法可用于处理网络结构和动力机制耦合这个问题。例如，我们可以通过最优化构建网络模型，其中需要优化的量与网络流量或搜索这样的动力机制相关。更加完善的方法在于修改适应性模型，使得适应性的值取决于一些动态参数。当动力机制继续作用时，适应性也会相应地发生改变；这便可以重组网络。其他策略也是可行的，所有这些都巩固了一个基本观点：多数情况下，当某个动力机制在网络结构中发生或与它耦合时，多数时候我们必须将基本图纳入考虑以充分理解正在发生的情形。

第九章 整个世界是否就是一张网？

据报道，量子力学的创始人之一保罗·狄拉克在谈及物理学家们在20世纪初的革命性发现时说："其余都是化学过程。"他的意思是，所有科学都能从物理学的基本原理中推导出来。很不幸，仅有少数几种情况可以由量子力学方程精确求解，它们还基本都是氢氦原子。像分子那样更为复杂的东西则必须以近似法或计算机模拟的方式处理。目前，除了少数宏观量子效应，基础物理学当前对我们理解生物、心理或社会而言相对无用。类似的错误还发生在遗传学中，人们错误地将DNA界定为能够解释人类所有特征、疾病和行为的决定因素。一般而言，基础科学成果的应用不应超出其真实的有效范围，人们应当认识到更专业的学科能够提供超出这一范围的更深刻洞见。网络科学也应避免落入被夸大的圈套。其整体性的视野，对差异明显的系统出人意料的相似性的揭示，还有当前对网络概念的文化上的迷恋，所有这一切使得我们很容易将网络科学视为"万用理论"。社会学家、工程师、生物学家和哲学家都曾警告世人注意从网络理论中提取的空洞概括。这类批评多数是合理的，但网络科学的成果不应被低估，其对未来各种发现的潜力也不应受到贬低。

网络科学的第一个主要限制是其缺乏大规模数据。社会科学中使用的调查问卷和访谈等方法十分昂贵且耗时，有时候还容易带有主观偏见。从信息技术（通话记录、电子邮件、社交网络、地理定位、无线射频识别芯片、健康数据、信用卡等等）处获取的数据为我们了解人们的社交关系提供了空前的便利，但这也带来一些问题。披萨送货员会接到许多电话，但多数来自客户而非其朋友："披萨送货员问题"表明，人们从信息技术数据（比如本例中的通话记录）中整理出相关信息并不容易。此外，还应提及的是，网络数据挖掘也制造了一些涉及隐私以及军事用途的伦理问题。

很多情况下，人们仅能获取部分信息。为了画出水生生物的食物网，生态学家会捕捉鱼类并检查其消化道：即便最杰出的科学家用这种方法也会错过一些物种联系。推导出蛋白质之间物质相互作用的基因方法既能产生假阳性也能产生假阴性结果。互联网和万维网图可以通过从某节点释放"探测器"以探索其周围各边的方式获取：足够数量

的路径能很好地呈现网络，但某些边可能从不会被探测器发现。

一旦数据可用，将其呈现在图中则不可避免会将其简化。网络方法的优势之一是其对拓扑的关注，但它忽略了元素的诸多具体特征。如果我们对这些特征感兴趣，图形近似法便显得不够用了。网络模型有时候被修改后可以包含这些特征，但并不总是如此。

当地理（即节点的物理位置）的重要性高过拓扑时，图示法也会不够用。例如，变电站、机场或火车站的位置明显与其连接的安排有关。此外，社交网络和食物网中的节点接近性则能决定建立某种关系的实际可能性。图示法可能遗漏的另一个元素是时间。例如，在性传播疾病中，在某人被他人感染之后与其建立关系明显不同于在这之前与其建立关系：关系建立的时机对疾病的传播至关重要。时机在大量网络中都很重要：例如，科学出版物按时间排序，科学家们只能引用那些之前发表的文献。

有时，识别节点和边绝非小事。我们很容易区分鹰和隼，却很难数得清生态系统中的细菌数量。为了避免低估“微小”物种的数量，生态学家通常会将营养物种里的生物聚集，这些生物共享同样的天敌以及相同的猎物。类似地，社会科学家则会聚集结构上等同的个人：具有相同联结数量和种类的人，比如家庭成员。类似的方法还被用于自主系统层面的互联网以及大脑区域网络等。这些过程必须以连贯的方式进行以获得一个合理的网络。边的定义更为复杂。某公司可能持有另一个公司的小部分或全部资本。两座机场可通过每天一次航班或每小时一次航班相互连接。在所有这些情况中都必须建立阈值，低于该阈值则被视为弱关系而不被记录。为连接加权或设置阈值会强烈影响所得网络的形状，因此必须建立在充足的理由上。

一旦将数据组织为图，就必须仔细解读其结果。网络科学有一整个学科分支都在专门研究可视化，即制造算法以便在纸张或计算机屏幕上合理安排节点和连接。然而，多数结论无法经由目测得出，而要对其进行数学分析。一些批评指出，并非所有复杂网络都具备异质度数分布（在任何情况下，这都绝非数学上的精确幂律）。的确，若网络并非异质，它们也会有趣，但公平地说，有趣的网络往往是异质的。完美的幂律并不重要：重要的是肥尾的出现，它揭示了枢纽节点的存在。将异质性解读为自组织的标记这一点已受到批评，人们指出许多网络中都存在着一定程度的规划，正如管理员对局域网的具体设计一样。但无疑地，互联网与其他许多网络一样并未经过大规模设计。因此，人们有理由认为，这些网络对随机性的偏离可被归因于自组织过程。此外，度数异质性仅是网络复杂性的其中一个标记。异质性还出现在许多其他特征中，比如中介性、集聚性和权重等，复杂性

同样也会出现在其他特征中，这一点与异质性不同：例如，模块和社区结构常常偏离随机性，并对网络顶层的动力机制产生强烈影响。

另外一个批评则指出，网络科学发现的只是不同系统和动力机制之间的模糊相似性，而非真实的普适类。后者是对应于相同基本数学定律（在特定细节被忽略的情况下）的不同现象群。当然，生物网络的具体特征与技术网络的具体特征完全不同，计算机蠕虫病毒的扩散与疾病的传染遵循着不同的规律。然而，网络理论为这些如此不同的结构和过程提供了一个共同的趋势和预测的框架。通常，如果系统足够大且相关现象的观察时间足够长，它们便会展现出十分相似的趋势。

网络科学已经在一些领域显示出了它的预测能力。它目前被应用于咨询行业，以帮助组织更好地开发各成员不同的技能；在公共卫生领域，它能预测和防止传染病的蔓延；在警察和军事系统中，它被用于追踪恐怖分子、罪犯以及反叛团伙；此外网络科学在其他一些领域也有应用。还有许多问题亟待解决，其中包括：制作更精细的模型以适配特定的网络和动力机制；寻找新的相关网络数据；深入挖掘拓扑以发现未被注意到的规律，并充分解释现有规律；描绘小型网络的特征并学习如何处理众网之网；将生物网络与演化范式更有效地连接；发现新的应用（比如在药物设计中）；以及有可能的话找到普适类。

伽利略·加利莱伊有一句名言如此说道：“哲学写在宇宙这本大书之中……书写语言为数学，字母则为三角形、圆形以及其他几何图形……”我们相信，尤其在如今这个复杂的当代世界里，我们需要网络这种“字母”。

注释

[1]译文引自朱生豪先生译本。——译注

[2]“pupil”还有“瞳孔”的意思。——编注

[3]美国无线运营商。——编注

[4]G.C.及后面的M.C.即本书的作者圭多· 卡尔达雷利与米凯莱· 卡坦扎罗。——编注

[5]此处指位于马萨诸塞州剑桥市的哈佛大学神学院。——编注